AD_____

Award Number: DAMD17-02-1-0634

TITLE: Spectral Analysis of Breast Cancer on Tissue Microarrays: Seeing Beyond Morphology

PRINCIPAL INVESTIGATOR: David L. Rimm, M.D., Ph.D.

CONTRACTING ORGANIZATION: Yale University School of Medicine
New Haven, CT 06520-8047

REPORT DATE: May 2004

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 2004	Annual (15 Apr 2003 - 14 Apr 2004)

4. TITLE AND SUBTITLE
Spectral Analysis of Breast Cancer on Tissue Microarrays: Seeing Beyond Morphology

5. FUNDING NUMBERS
DAMD17-02-1-0634

6. AUTHOR(S)

David L. Rimm, M.D., Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Yale University School of Medicine
New Haven, CT 06520-8047

E-Mail: david.rimm@yale.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Original contains color plates: ALL DTIC reproductions will be in black and white

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

Spectral imaging is a mechanism of tissue examination that combines spectral (color) with spatial information. The Varispec™ device fractionates the light into 10 nm bands and then collects images containing a full spectral profile for each pixel. The resulting spectral profile is then analyzed using a genetic algorithm based software package called GENIE. GENIE allows combination of spectral and spatial information for binary inclusion into user defined classes. This IDEA grant proposed examination of the spectral signatures of breast cancer to distinguish benign cells from malignant cells allowing diagnosis, classification, and possibly prediction of outcomes in breast cancer. The first stages of the grant analyzed tissue microarray spots, to assess our ability to distinguish benign from malignant tissue using very "easy" cases. This has been completed with good sensitivity and specificity and more challenging specimens are now being assessed. The second part of the proposal focused on application of this technology to cytology specimens. Due to some difficulties in obtaining optimal material for breast cytology, urine cytology specimens have been used as a model system. Using this model, we have been able to distinguish benign from malignant cells with high accuracy and even shown good accuracy in adjudication of "atypical" specimens.

14. SUBJECT TERMS
Breast Cancer

15. NUMBER OF PAGES
39

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Introduction

Conventional analysis of breast cancer specimens has largely been based on the microscopic appearance of the tumor. Surprisingly, the microscopic and molecular analyses of tissue have ignored color, a potentially tremendous source of information. Preliminary data suggests that the information content of the spectra of tissue is as high, or higher than that obtained from conventional spatial morphology. Recently, the combination of new optical technologies (spectral imaging) and vastly improved computer power has evolved such that quantitative spectral analysis can be done on each pixel of a complex histologic image. The purpose of this project was to test the hypothesis that spectral analysis will provide diagnostic and prognostic information beyond that attainable from conventional morphology using the same starting material, a stained histology or cytology slide. To test that hypothesis we proposed a three-fold approach, First, we will determine the ability of spectral analysis to distinguish benign from malignant breast tumors. Secondly, we will determine if spectral information can segment patient cohorts based on outcome (in a manner analogous to the way conventional morphology uses histologic and nuclear grade). Finally, we will assess the whether the spectral signatures can be used in a broader fashion to aid diagnosis in cytologic specimens.

Body

The original approved statement of work was as follows:

Tasks/Aims:
Aim 1: To use spectral analysis to classify benign from malignant in breast tissue specimens
Aim 2: To use spectral analysis to attempt to stratify breast cancers with respect to outcome in a manner comparable to histologic/nuclear grade, clinical stage, or prognostic marker results.
Aim 3. To use spectral signatures to classify cytologic breast fine needle aspiration specimens.

Year 1:
1. Construct benign/malignant tissue array from existing tissue collections using approximately 250 cases of breast cancer and associated normal tissue. Assemble associated databases.
2. Hire fellow and train on spectral analysis software.
3. Begin pilot analysis of breast tissue, defining benign/malignant spectral signatures.
4. Select and/or prepare cytologic FNA specimens for Aim 3 cohort.

Year 2:
1. Collect and analyze data on benign/malignant array and define spectral signatures.
2. Completion of preliminary work and publication of first description of methods for spectral-based pathologic analysis.
3. Begin analysis of the classification potential of spectral signatures by collection of spectral data from arrays where the stage, grade, and outcome information is used in selection of machine training regions. This will probably include integration and further training on new software (as it is developed)
4. Optimization/standardization of staining protocols for cytology specimens.
5. Collection of spectral data and data analysis of control cytologic specimens.

Year 3:
1. Completion and publication of first efforts on spectral classification.
2. Continuation of analysis of the classifying capacity of spectral signatures by optimization of the information used in selection of machine training regions. This may include integration of new software as it is developed.
3. Application of spectral signature to cytologic specimens to attempt to stratify on the basis of benign vs malignant, but then also to classify "atypical" cases based on their spectral profile. This may also require further training on cytology-specific modifications of the software.

To date we have completed nearly all of the tasks targeted for years one and two and made some progress on other tasks. Specifically, benign and malignant breast cancer tissue microarrays have been constructed and the relevant clinical follow-up information has been collected. Raj Jaganath learned to use the Varispec™ device and software and collected image stacks on a 20 benign and malignant spots. He was trained and assisted in this effort by Dr. Richard Levenson, a key consultant to the project. The images were reviewed and annotated by Tolgay Ocal, a collaborating expert breast pathologist, then sent to Neal Harvey at Los Alamos National Labs for Genie-based software analysis. GENIE is Unix-based software produced at Los Alamos, based on the genetic algorithm concept using both spatial and spectral data as the computational basis. The analysis software is not yet usable by general pathologist. However, Dr. Harvey used the annotations of Dr. Ocal to define normal from malignant and then constructed training sets based on the spectral profiles of a series of spots. Then a second series of spots was selected as a test set. The result was that over 87% of all cancerous nuclei pixels were correctly identified while less than 7% of

normal tissue was incorrectly labeled as cancer. For images that contained only normal tissue, on average, GENIE incorrectly labeled less than 1% of pixels as cancer. Thus, although this is preliminary data, it is very promising. It has been presented at the SPIE Biomedical Imaging Conference of 2003 and published in the meeting proceedings. The publication is included in the appendix. In follow up to this study, Dr. Angeletti has collected images from a much larger series of benign and malignant spots from these arrays to extend this study. This study has taken longer that anticipated and is still in progress.

Toward to goals of aim 3, Raj has prepared a series of 80 breast FNA specimens for analysis by prepping and staining the specimens in identical manner. Images have been collected from this series of FNAs and analysis of the images is underway.

As of July 1, 2003, Cesar Angeletti was joined this effort on a full time basis replacing Raj Jaganath who graduated. Shortly after he began in the lab, Cesar visited our collaborator, Neal Harvey at Los Alamos National Labs and learned to use the GENIE program described above. He then began analysis of a series of experiments on tissue and cytology specimens at Yale. Although he focused on breast cancer, for the purposes of testing of the GENIE system on cytology specimens, he has begun a series of studies on urine specimens. He chose these specimens since they are much more plentiful than breast FNAs and provided an easy-to-use test platform in preparation for the more complex analysis required for breast FNAs.

Using these specimens, he has directly addressed the staining issues raised in SOW point 4 for year 2. We have evaluated the effect of staining on spectral properties, both across years and batches at our own institution as well as between institutions, in a collaboration forged with Andrew Fisher at the University of Massachusetts in Worcester. We selected and process of hundreds of cells from three data sets: 1) Yale specimens from 1996-1997, 2) Yale specimens from 1998-1999 and 3) UMass specimens from 2003-2004. Using a GENIE derived chromosome to define malignant cells, based on a fraction of the '96-'97 set, we found a sensitivities and specificities in the 85-95% range for correct assignment of malignant cells, independent of year of acquisition or institution. (see table 1) This result suggests that we should be able to use to standard Pap stain for breast cancer specimens and that the information content of those specimens should be sufficient for similar spectral cytologic analysis.

Table 1

Data Set	N Benign	N Malignant	Sensitivity	Specificity
Yale 96-97	308	122	91%	95%
Yale 98-99	190	178	87%	96%
UMASS 03-04	121	40	85%	97%

The next step was the analysis of "atypical" specimens. Both Breast and urinary cytology suffer from the fact that in 25% or more of the specimens, the morphologic information is insufficient for a definitive diagnosis. The result is that the pathologist calls these cases "atypical" which is not very helpful in the next steps of patient management. One of the goals of this technology is to eliminate the "atypical" diagnosis by the addition of spectral information. Again, this was piloted in urine specimens due to the difficulty in obtaining breast FNAs. Here we selected a series of cases where the cytopathologist made the diagnosis of atypical and the urologist then decided to biopsy the patient. We used the histological biopsy result as the criterion standard and constructed a training set of atypical urine specimens adjudicated by cystoscopic biopsy. We found we were able to generate a GENIE chromosome that could predict the biopsy result with 70-80% accuracy. This data has now been synthesized in a manuscript that has been submitted to Nature Methods (attached as an appendix).

Key Research Accomplishments:

1. Completion of Initial Breast Tissue Microarrays for Malignant vs Normal and Outcome-based analysis

2. Completion of Spectral Image stack acquisition for Malignant vs Normal series.

3. Completion of Analysis of Malignant vs Normal series and construction of training and out-of-training set analyses

4. Collection of cases for Breast FNA studies

5. Completion of standardization and testing of cytology specimen and staining parameters.

6. Completion of a series of urine specimens as a model for the breast cancer studies

7. Adjudication of atypical cytology specimens by the combination of spectral and spatial information using GENIE algorithms

Reportable outcomes:

An abstract, which was subsequent published in the SPIE proceedings describes the combination of spectral and spatial analysis showing good classifying ability to distinguish normal breast tissue from malignant tissue (see appendix).

A paper has been published in the journal Cancer Cytopathology related to the spectral properties of cytology specimens using urine specimens as a model system, a second manuscript is submitted using similar material with the addition of the GENIE system for analysis (see appendix).

Conclusions:

Preliminary results suggest there is sufficient information attainable from the combination of spectral and spatial data, using genetic algorithms, to classify malignancy in breast cancer. We have also now tested the system on Papanicoulaou stained cytology specimens. We have shown that the information obtained is from this material is robust and independent of stain batch or even institution. We are now facing the more difficult challenges of distinguishing benign lesions from malignant lesions and correlation of spectral/spatial features with tumor behavior.

References:

See original proposal and reference sections of appended manuscripts.

Appendix Contents:

1. Harvey, N.R., Levenson, R.M., Rimm, D.L. (2003) Investigation of Automated Feature Extraction Techniques for Applications in Cancer Detection from Multispectral Histopathology Images. Proc. of SPIE, 5032: 557-566

2. Jaganath, R., Angeletti, C., Levenson, R., and Rimm, D.L. (2004) Diagnostic Classification of Urothelial Cells in Urine Cytology Specimens Using Exclusively Spectral Information. Cancer Cytopathology. Cancer 102:186-91.

3. Angeletti, C., Harvey, N., Khomitch, V., Fischer, A., Richard Levenson, R. & Rimm, D.L. (2005) Spectral-Spatial Analysis of Urine Cytology Specimens. Submitted to Nature Methods

Investigation of Automated Feature Extraction Techniques for Applications in Cancer Detection from Multispectral Histopathology Images

Neal R. Harvey*[a], Richard M. Levenson[b], David L. Rimm[c]

[a]NIS-2, Los Alamos National Laboratory, Los Alamos, NM, 87545;

[b]Cambridge Research and Instrumentation Inc., 35-B Cabot Road Woburn, MA 01801;

[c]Dept. of Pathology, Yale University School of Medicine, 310 Cedar St., New Haven, CT 06520

ABSTRACT

Recent developments in imaging technology mean that it is now possible to obtain high-resolution histological image data at multiple wavelengths. This allows pathologists to image specimens over a full spectrum, thereby revealing (often subtle) distinctions between different types of tissue. With this type of data, the spectral content of the specimens, combined with quantitative spatial feature characterization may make it possible not only to identify the presence of an abnormality, but also to classify it accurately. However, such are the quantities and complexities of these data, that without new automated techniques to assist in the data analysis, the information contained in the data will remain inaccessible to those who need it. We investigate the application of a recently developed system for the automated analysis of multi-/hyper-spectral satellite image data to the problem of cancer detection from multispectral histopathology image data. The system provides a means for a human expert to provide training data simply by highlighting regions in an image using a computer mouse. Application of these feature extraction techniques to examples of both training and out-of-training-sample data demonstrate that these, as yet unoptimized, techniques already show promise in the discrimination between benign and malignant cells from a variety of samples.

Keywords: multispectral, histopathology, classification, cancer, machine learning

1. INTRODUCTION

In the field of pathology, accuracy in tissue diagnosis is essential to ensure that patients receive the most appropriate, most cost-effective and least toxic therapies. At present, the state of the art for the determination of a pathological diagnosis relies on manual, morphology based analysis of tissue sections, a method largely unchanged since the nineteenth century. Relying largely upon visual pattern recognition of tissue samples, the entire process is subjective, somewhat irreproducible and inefficient in extracting all the information contained in the specimen, especially as related to prognosis and therapy guidance. Recent advances in optical technologies, coupled with improved computer power, mean that it is now possible to extract information beyond the capabilities of the human visual system. We can extend beyond the limitations of the human eye's acuity and the visible spectrum and obtain high-resolution histological image data at multiple wavelengths. These data have the potential for revealing (often subtle) distinctions between different types of tissue that could be useful in determining objective, reproducible disease-classifying information. The spectral content by itself contains a great deal of information, whose value increases greatly when it is combined with the spatial information available. Unfortunately, such are the quantities and complexities of these data, that without new automated techniques to assist in the data analysis, the useful information contained in the data may remain largely inaccessible. Integration of the spectral and spatial information contained in these images using sophisticated but robust statistical techniques should make it possible to obtain disease classifications that are more accurate, objective and reproducible than is possible with existing manual methods.

Here we describe preliminary experiments in which we investigate the application of a recently developed system for the automated analysis of multi-/hyper-spectral satellite and aerial image data to the problem of cancer detection from multispectral histopathology image data. The system, known as GENIE, was originally developed for the military and intelligence community, to provide a means to develop automated feature extraction tools for multi-

* harve@lanl.gov; phone +1 505 667 9077; fax +1 505 665 4414

and hyper-spectral aerial and satellite imagery. The reason for GENIE's development is that while there exist highly-skilled image analysts who are expert at identifying features of interest from complex image data sets, they are limited in number and, being human, have limited capabilities: they have a limited spectral capability (3 channels) and rate at which they can analyze imagery. So, in order to go beyond these limitations, there is a need to develop systems in which the power of modern computers and machine-learning techniques can be brought to bear. Although human analysts are extremely good at finding features of interest within imagery, they are not so good at describing exactly how they are able to do this, and hence, hand-coding algorithms designed for specific tasks is a difficult and often long and expensive process. Thus, we have developed a system whereby a human expert can teach a computer to create algorithms to perform these functions, via a simple graphical user interface, in which a human provides training data to the computer by simply highlighting examples of the features of interest on a computer screen.

One can make certain comparisons between the military and intelligence community and the medical (specfically, pathology) community. They both have a great deal of complex, high-dimensional data (multi- and hyper-spectral satellite and aerial imagery vs multi-spectral histopathology imagery) that they wish to analyze. They both wish to find features of interest within complex backgrounds (e.g. military targets vs cancer cells) and they both have human experts available who are highly skilled at identifying these features (image analysts vs pathologists), but who have limitations with regard to the complexity and quantity of the data which they can analyze. Bearing these similarities in mind, it is not unreasonable to investigate the application of a software system originally developed to address remote-sensing problems to a set of problems in the medical arena.

2. SPECTRAL IMAGING

Spectral imaging microscopy represents a technological advance over visual or RGB-camera-based analyses, providing images at multiple wavelengths and generating precise optical spectra at every pixel. These rich data sets have applications in surgical pathology, multicolor fluorescence and immunohistochemistry. There now exists a variety of technologies for use in combination with microscopy, including tunable filters, Fourier-transform interferometry, line-scanning prism or gratings-based devices, computed tomography, and others based on polarization effects. Mathematical approaches to these complex data sets may then be used to extract maximum possible information from the resulting data.

In the experiments described here, a VariSpec(tm) liquid crystal tunable filter devices (CRI, Inc.)[14] was used. This device can transmit in a number of wavelength ranges (e.g., 400-720 nm or 850-1800 nm with bandwidths typically in the 7 to 20-nm range, although bandwidths as narrow as 0.1 nm have been achieved).

3. AUTOMATED IMAGE ANALYSIS: OVERVIEW OF THE GENIE SYSTEM

The details of GENIE's algorithmic structure have been described previously in the literature,[1–7] so, in the interests of brevity, we provide only a brief overview of our system.

Our particular interest is the pixel-by-pixel classification of multi-spectral images, not only to locate and identify but also to delineate particular features of interest. For the experiments described here, we are interested in distinguishing cancerous (malignant) cells against the background (which includes normal benign cells). Due to the quantities and complexities of the multispectral data with which we are working, the hand-coding of suitable feature-detection algorithms is impractical. We therefore use a supervised learning approach that can, using only a few hand-classified training images, generate image processing pipelines that are capable of distinguishing features of interest from the background. We remark that our approach here is to consider the two-class problem: although many classification applications require the segmentation of an image into a larger number of distinct classes, for our particular problem, our main interest is the simpler problem of identifying a single class (cancer) against a background of "other" classes. GENIE does possess the capability for performing multiple-class classification,[8] but here we did not make use of that functionality.

GENIE employs a classic evolutionary paradigm: a population is maintained of candidate solutions (*chromosomes*), each composed of interchangeable parts (*genes*), and each assessed and assigned a scalar fitness value, based on how well it performs the desired task. After fitness determination, the evolutionary operators of selection, crossover and mutation are applied to the population and the entire process of fitness evaluation, selection, crossover and mutation is iterated until some stopping condition is satisfied.

3.1. Environment

The environment for each individual in the population consists of *data* planes, each of these planes corresponding to a separate spectral channel in the original multi-spectral image, together with a *weight* plane and a *feature* plane. The weight plane identifies those pixels to be used in training – these are all the pixels for which the analyst has provided a class label. The actual delineation of separate feature/class pixels is given by the feature plane.

3.2. Chromosomes and Genes

Each individual *chromosome* in the population consists of a fixed-length string of *genes*. Each gene in GENIE corresponds to a primitive image processing operation. Therefore the entire chromosome describes an algorithm consisting of a sequence of primitive image processing operations.

Each gene used in GENIE takes one or more distinct image planes as input, and produces one or more image planes as output. Input can be taken from any of the data planes in the training data image cube. Output is written to any of a small number of *scratch planes* — temporary workspaces where an image plane can be stored. Genes can also take input from scratch planes, but only if that scratch plane has been written to by another gene earlier in the chromosome sequence.

Our "gene pool" is composed of a set of primitive image processing operators which we consider useful. These include spectral, spatial, logical and thresholding operators.

3.3. Backends

Final classification requires that the algorithm produce a single (discrete) scalar output plane, which identifies, for every pixel, the class to which it has been assigned. We have found it advantageous to adopt a hybrid approach which applies a conventional supervised classifier to a (sub)set of scratch and data planes to produce the final output plane.

To do this, we first select a subset of the scratch and data planes to be *answer planes*. The conventional supervised classifier "backend" uses the answer planes as input and produces a final output classification plane; in principle, we can use any supervised classification technique as the backend, but for the experiments reported here, we used the Fisher linear discriminant[9] as the backend.

3.4. Fitness Evaluation

The fitness of a candidate solution is given by the degree of agreement between the final classification output plane and the training data. It is based on a simple ratio of the total number of incorrectly classified training pixels over all classes to the total number of training pixels over all classes. If we denote the detection rate (fraction of "true" pixels classified correctly) as R_d and the false alarm rate (fraction of "false" pixels classified incorrectly) as R_f, then the fitness F of a candidate solution is given by

$$F = 500(R_d + (1 - R_f)). \tag{1}$$

Thus, a fitness of 1000 indicates a perfect classification result. This fitness score gives equal weighting to type I (true pixel incorrectly labelled as false) and type II (false pixel incorrectly labelled as true) errors. Note a fitness score of 500 can be trivially achieved with a classifier that identifies all pixels as true (or all pixels as false).

4. EXPERIMENTS: CANCER DETECTION

4.1. Tasks

We set GENIE the task of detecting cancerous nuclei in multispectral breast tissue image data. Thus we have a classification problem with two classes: (1) cancerous nuclei and (2) everything else. Therefore, GENIE was given the task of searching for algorithms that would be able to label each pixel within an image as belonging to one or other of these two classes. While our approach here was to consider the two-class problem, we are aware that other applications might require the segmentation of an image into a larger number of distinct classes. In fact, GENIE is capable of addressing multiple-class problems.[8] However, for this study, we only consider the simpler problem of identifying a single class against a background of "other" classes.

4.2. Multispectral data

The construction of tissue microarrays (TMAs) has been previously described and recently reviewed.[10–13] Briefly, formalin-fixed, paraffin-embedded tissue blocks containing breast cancer were retrieved from the archives of the Yale University Department of Pathology. Areas of invasive carcinoma were identified on corresponding hematoxylin-eosin stained slides and the tissue blocks were cored and transferred to a recipient "master" block using a Tissue Microarrayer (Beecher Instruments, Silver Spring, MD). Each core is 0.6 mm wide, spaced 0.7–0.8 mm apart. After cutting of the recipient block and transfer with an adhesive tape to coated slides for subsequent UV cross-linkage (Instrumedics, Inc, Hackensack, NJ), the slides were dipped in a layer of paraffin in order to prevent oxidation (24). Slides were stained with hematoxylin and eosin, were evaluated for quality of the section and then selected for spectral imaging analysis. For these experiments, examples of both breast cancer and normal tissue were selected.

Images were collected at 10 nm intervals between 420 nm–700 nm using a CRI (Cambridge Research Instruments[15]) VariSpec filter, CRI PanKroma acquisition software, a light microscope, and a QImaging Retiga megapixel digital monochrome camera. The process is semi-automated. The image on the CCD is brought into focus while the tunable filter is tuned to 550 nm (a high-contrast part of the spectrum for H & E samples). An autoexpose function then steps the filter through the spectral range, calculating exposure times wavelength-by-wavelength that will cause the brightest pixels to nearly fill their dynamic range (250 counts for an 8-bit, 256-level sensor). Using these exposure times, a stack of images is automatically collected, with the computer tuning the filter and acquiring an image at every wavelength step, resulting in stacks of 29 images for each sample (tissue microarray dot). To remove optical irregularities in the image train (dust on the CCD window for example) and also some variations in intensity across the liquid crystal filter, the images are flat-fielded by dividing (and normalizing for intensity) each plane of the sample image by the corresponding plane of a white stack obtained from a clear area on the same slide. The image stack, consisting of a series of tif images sequentially numbered, is converted into a single ENVI-format data file with separate header, and transferred to Los Alamos via ftp.

4.3. Training data

In order to provide training data, several images were selected, some containing a mixture of both cancerous and normal tissue and some containing only normal, healthy tissue. Of these images, sub-regions were selected that contained suitably representative samples of pixels from both classes: (1) cancerous nuclei and (2) "everything else". For the cancerous nuclei training samples, regions that had a high density of cancerous nuclei were selected. For the "everything else" training samples, regions were selected that had combinations of the kinds of features that are present in that somewhat-broad class. We were careful to select some regions of normal tissue that contained a high density of normal, healthy nuclei, in order to provide some training data samples that could assist GENIE in evolving an algorithm able to successfully disambiguate cancerous from healthy nuclei.

Fig. 1 shows examples of the original image data and the associated training data (labels) provided by the expert. Fig. 1 (a) shows a true color image of one of the images obtained for breast tissue containing cancer. Fig. 1 (b) shows the training data provided by the expert for the data shown in Fig. 1 (a). Pixels labelled as containing cancer are colored green and pixels labelled as normal are colored red. The training data p(red and green) image has been overlaid onto a gray-scale representation of the true-color image shown in Fig. 1 (a). The region enclosing only those pixels in the image used for training is shown by the bounding box. Fig. 1 (c) shows a true color image of one of the images obtained for breast tissue containing only normal tissue. Fig. 1 (d) shows the training data provided by the expert for the data shown in Fig. 1 (c). As with Fig. 1 (b), pixels labelled as containing cancer are colored green and pixels labelled as normal are colored red (note that there are no green pixels in this image). The training data image has been overlaid onto a gray-scale representation of the true-color image shown in Fig. 1 (c), and the region enclosing only those pixels in the image used for training is shown by the bounding box.

5. RESULTS

Fig. 2 shows the results of applying the classification algorithm found by GENIE during its training, to some data. Fig. 2 shows the classification results of applying the algorithm to the data shown in Fig. 1 (a). The pixels labelled by the algorithm as cancer are colored green and those labelled as normal are colored red. The resulting classification (red and green) image has been overlaid onto a gray-scale image of the original data, just as for the training data shown in Fig. 1 (b). Fig. 2 (b) shows a true-color image of a data set containing cancerous and normal tissue that was not seen during training. Fig. 2 (c) shows a true-color image of a data set containing only normal tissue that

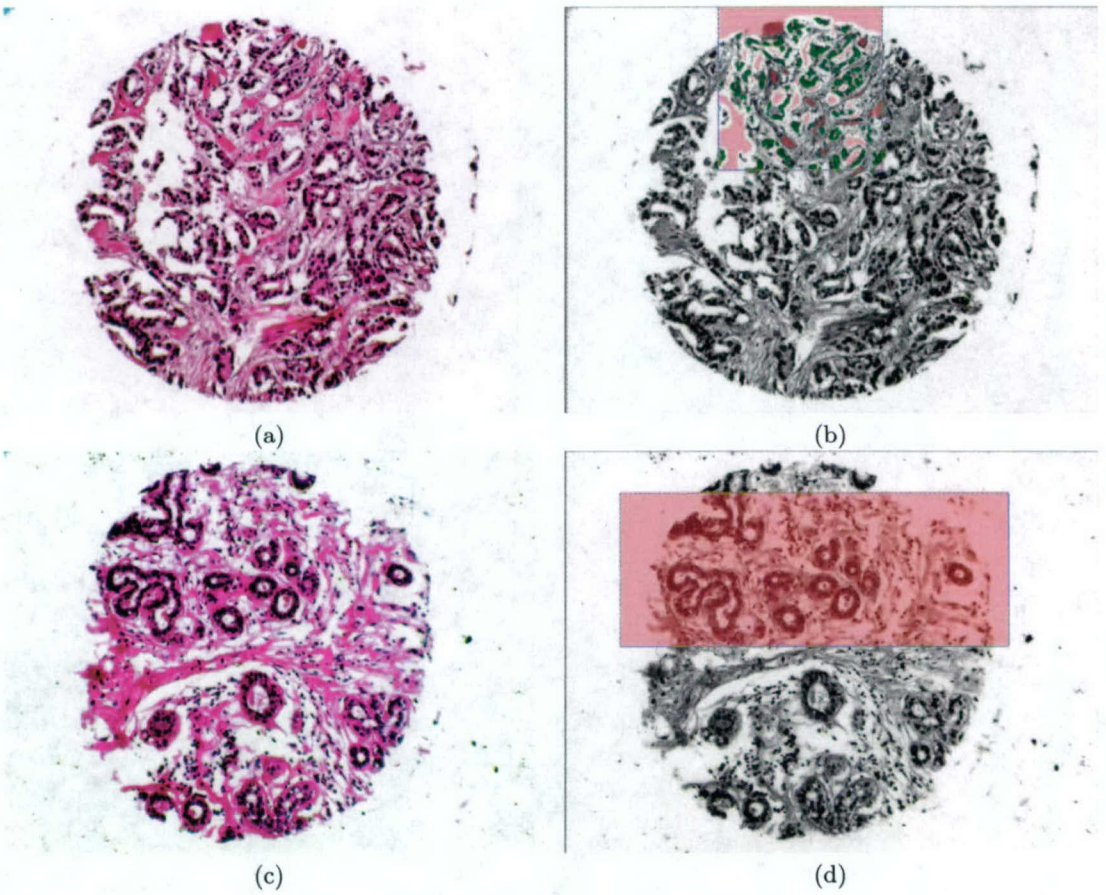

Figure 1. Breast: (a) True color image of one of the data sets obtained from breast tissue containing cancer; (b) Training data provided from this image: Green = Feature (i.e. Cancer), Red = Not Feature (i.e. Non-Cancer); (c) True color image of one of the data sets obtained from breast tissue containing only healthy (non-cancerous) tissue; (d) Training data provided from this image: Notice there are no Green pixels, due to there being no cancer in the image.

was not seen during training. Fig. 2 (d) shows the classification results of applying the algorithm to the data shown in Fig. 1 (c). As before, the pixels labelled by the algorithm as cancer are colored green and those labelled as normal are colored red, and the resulting classification image has been overlaid onto a gray-scale image of the original data.

Table 1 shows the performance of the algorithm found by GENIE during training, as relates to the training data and the entire images, from which the training data was extracted. Column 1 shows the image name. Column 2 shows the number of pixels labelled as cancer that were provided in the training data for each image. Column 3 shows the number of pixels labelled as non-cancer (normal) that were provided in the training data for each image. Column 4 shows the detection rate, DR, (percentage of pixels labelled as cancer in the training data that were labelled correctly as cancer by the algorithm found by GENIE during training) for each image in the training data set. Column 5 shows the false-alarm rate, FAR (percentage of pixels labelled as normal in the training data that were incorrectly labelled as cancer by the algorithm found by GENIE during training) for each image in the training data set. Column 6 shows the total number of pixels labelled as cancer by GENIE's algorithm for the entire image from which the training data was extracted.

Table 2 shows the performance of the algorithm found by GENIE during training, as relates to some testing data - i.e. some image data which was not seen during training (out-of-training-sample data). For these images, in order to be able to assess GENIE's performance in a quantitative manner, an expert provided ground truth for regions in these images, in a similar manner to that provided for the data used in training. Column 1 shows the image name.

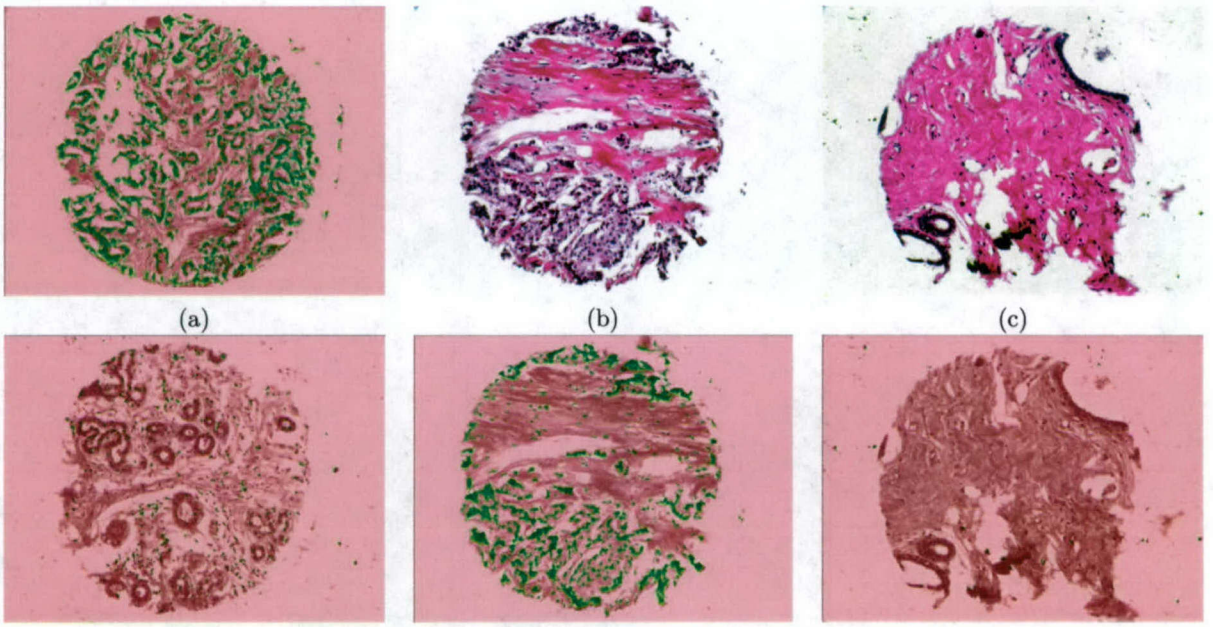

Figure 2. GENIE: Breast (a) Output of GENIE-derived classification algorithm found during training, applied to raw multispectral data shown in Fig. 1 (a); (b) True color image of one of the data sets obtained from breast tissue containing cancer, but not used during training; (c) True color image of one of the data sets obtained from breast tissue containing only healthy (non-cancerous) tissue, but not used during training; (d) Output of GENIE-derived classification algorithm applied to raw multispectral data shown in Fig. 1 (c); (e) Output of GENIE-derived classification algorithm found during training, applied to raw multispectral data shown directly above in (b); (f) Output of GENIE-derived classification algorithm found during training, applied to raw multispectral data shown directly above in (c)

Column 2 shows the number of pixels labelled as cancer by the expert for each image. Column 3 shows the number of pixels labelled as non-cancer (normal) for each image. Column 4 shows the detection rate (DR) for each image in the testing data set. Column 5 shows the false-alarm rate (FAR) for each image in the testing data set. Column 6 shows the total number of pixels labelled as cancer by GENIE's algorithm for the entire image, not just the region labelled by the expert.

Table 2 shows the performance of the algorithm found by GENIE during training, as relates to some testing data - i.e. some image data which was not seen during tyraining (out-of-training-sample data), but for which we don't have expert-provided ground truth. While we don't have expert-provided ground-truth on a pixel-by-pixel basis for these images, we do know, for each image, whether it contains some cancer or whether the image has only normal tissue. Thus, for these images we only provide the total number of pixels labelled as cancer by GENIE's algorithm for the entire image.

6. DISCUSSION

It can be seen, both from the images shown in Fig. 2 and in Tables 1 – 3, that GENIE was able to evolve an algorithm capable of doing a good job of discriminating cancer from non-cancer in the multispectral images used in these experiments. For the training data, for the images that contained both cancerous and non-cancerous (normal) tissue, GENIE was, on average, able to detect over 87% of all cancerous nuclei pixels and only incorrectly labelled less than 7% of normal tissue as cancer. For images that contained only normal tissue, on average, GENIE incorrectly labelled less than 1% of pixels as cancer. For testing data, for which an expert had provided ground-truth, for images that contained a mixture of both cancerous and normal tissue, GENIE, on average, was able to correctly label more

Table 1. Performance of the GENIE-derived classification algorithm found during training applied to training-sample data

Image Name	# Labelled Cancer Pixels (Training)	# Labelled Non-Cancer Pixels (Training)	DR (%)	FAR (%)	Total # Pixels Labelled as Cancer in Result Image
C1-15	25230	147657	70.32	0.17	60742
C2-12	27422	134501	94.72	18.63	138568
C2-9	14871	34187	97.28	0.25	215508
Average	22508	105448	87.44	6.35	138273
N1-8	0	132880	–	0.21	4408
N2-9	0	204768	–	0.21	554
N1-1	0	243120	–	0.36	1305
N4-4	0	335616	–	2.31	18318
Average	0	229096	–	0.77	6146

Table 2. Performance of the GENIE-derived classification algorithm found during training applied to out-of-training-sample data, for which an expert had provided labels, in order to determine out-of-sample performance

Image Name	# Labelled Cancer Pixels (Testing)	# Labelled Non-Cancer Pixels (Testing)	DR (%)	FAR (%)	Total # Pixels Labelled as Cancer in Result
C1-2	12357	83286	48.04	7.85	94274
C2-14	7992	27960	93.23	29.61	300216
C4-9	3880	54773	96.89	18.76	205827
C5-8	4006	73325	90.84	6.99	139858
Average	7059	59836	82.25	15.80	185044
N1-2	0	1.198×10^6	–	0.15	1746
N2-7	0	1.198×10^6	–	0.66	7938
N3-9	0	1.198×10^6	–	0.64	7708
N4-5	0	1.198×10^6	–	0.14	1640
Average	0	1.198×10^6	–	0.40	4758

than 82% of cancerous nuclei pixels and labelled less than 16% of normal tissue incorrectly as cancer. For images that contained only normal tissue, on average, GENIE incorrectly labelled less than 0.5% of pixels as cancer.

It should be noted that the non-nuclei, connective tissue surrounding the cancerous nuclei in the cancer-containing samples is, in fact, not normal tissue. It has its own deviation from normal. It is interesting to note that the algorithm evolved by GENIE labelled this tissue as cancerous. This is hardly surprising. The training data provided from the cancer-containing samples consisted of mostly pixels from cancerous nuclei, with very few samples from the surrounding stroma. However, there were plenty of training samples taken from normal, healthy stroma. Thus, with training samples provided for two classes: malignant nuclei and normal, healthy "everything else", it is understandable that malignant stroma would be significantly different from the training data samples provided for the normal healthy tissue, and would thus be classified into the other "cancerous nuclei" class.

While there was a drop in GENIE's performance, from training data to testing data, for images that contained both cancerous and normal tissue, with the average detection rate going from 87% to 82% and average false-alarm rate going from 7% to 16%, there was actually an improvement in performance, from training data to testing data, for images that contained only normal tissue, with the average false-alarm rate going from 3% to 0.4%.

In general, the algorithm discovered by GENIE does a very good job of discriminating cancer versus normal tissue, both for the data provided in training and for the out-of-training-sample data. There is a large difference (orders of magnitude) between the numbers of pixels classified as being cancer in those images containing cancer

Table 3. Performance of the GENIE-derived classification algorithm found during training applied to out-of-training-sample data, for which no labels had been provided

Image Name	Total # Pixels Labelled as Cancer in Result
C2-5	92385
C2-7	272119
C3-2	169649
C3-4	196517
C5-10	183402
Average	182814
N2-2	292
N2-4	1509
N3-4	686
N3-6	5348
N4-9	4107
Average	2388

compared to those images containing only normal, healthy tissue.

6.1. Further work

GENIE, as it currently stands, despite the promising results shown here, needs much modification in order to be made more generally useful for real applications in pathology. The present suite of operators that make up GENIE's "gene pool" are essentially those which were provided for remote-sensing applications. These operators are not necessarily the most appropriate for the field of pathology. A more targeted group of operators developed from those already developed for such applications in pathology and described in the literature[16,17] would be a good start. In addition, GENIE's current mode of operation, in which the classification is performed on a pixel-by-pixel basis is not ideal. Moving to a higher-level, more object-based classification methodology, would be a better approach. Going even further, beyond providing a simple binary classification indicating the presence or absence of cancer and providing a more detailed classification, such as cancer grade is an additional goal. The other area that needs work is to improve the time taken for training. At present, depending on the amount of training data provided and the complexity of the algorithm space GENIE is set the task of searching, it can take several hours to perform a training run. We aim to be able to reduce this training time to minutes. Our approaches to achieving this goal include parallelisation of the genetic algorithm,[3] implementation of image processing operators in hardware (via FPGAs[18]) and investigation of better, and more efficient search and classification methodologies.[19]

Further work also needs to be undertaken towards a proper validation of the approach, using a far greater volume of data than used in these experiments.

7. CONCLUSIONS

We have shown preliminary investigations into the application of a system originally developed for the automated analysis of satellite image data to the problem of cancer detection from histopathology image data. The results of this work shows great promise, but leaves many questions yet to be answered, and much work to be done.

REFERENCES

1. S.P. Brumby, J. Theiler, S.J. Perkins, N.R. Harvey, J.J. Szymanski, J.J. Bloch and M. Mitchell, "Investigation of Image Feature Extraction by a Genetic Algorithm", in *Proc. SPIE 3812* pp. 24–31 (1999).
2. J. Theiler, N.R. Harvey, S.P. Brumby, J.J. Szymanski, S. Alferink, S. Perkins, R. Porter and J.J. Bloch, "Evolving Retrieval Algorithms with a Genetic Programming Scheme", in *Proc. SPIE 3753*, pp. 416–425 (1999).

3. N.R. Harvey, S.P. Brumby, S.J. Perkins, R.B. Porter, J. Theiler, A.C. Young, J.J. Szymanski, and J.J. Bloch, "Parallel evolution of image processing tools for multispectral imagery", in *Proc. SPIE 4132*, pp.72-82, 2000.
4. S.P. Brumby, N.R. Harvey, S. Perkins, R.B. Porter, J.J. Szymanski, J. Theiler and J.J. Bloch, "A genetic algorithm for combining new and existing image processing tools for multispectral imagery", in *Proc. SPIE 4099*, pp. 480–490, (2000).
5. N.R. Harvey, S. Perkins, S.P. Brumby, J. Theiler, R.B. Porter, A.C. Young, A.K. Varghese, J.J. Szymanski and J. Bloch, "Finding golf courses: The ultra high tech approach", in *Evolutionary Image Analysis, Signal Processing and Telecommunications*, Poli, et al. Springer-Verlag (2000).
6. A.B. Davis, S.P. Brumby, N.R. Harvey, K. Lewis Hirsch, and C.A. Rohde, "Genetic refinement of cloud-masking algorithms for the multi-spectral thermal imager (MTI)" in *Proc. IGARSS 2001*, Sydney, Australia, 9-13 July 2001.
7. N.R. Harvey, J. Theiler, S.P. Brumby, S. Perkins, J.J. Szymanski, J.J. Bloch, R.B. Porter, M. Galassi, and A.C. Young, "Comparison of GENIE and Conventional Supervised Classifiers for Multispectral Image Feature Extraction", in *IEEE Trans. Geoscience and Remote Sensing*, 40:2, (2002), pp. 393–404.
8. N.R. Harvey, J. Theiler, L. Balick, P. Pope, J.J. Szymanski, S.J. Perkins, R.B. Porter, S.P. Brumby, J.J. Bloch, N.A. David, M. Galassi, "Automated Simultaneous Multiple Feature Classification of MTI Data", in *Proc. SPIE 4725*, 99. 346–356 (2002).
9. C.M. Bishop, *Neural Networks for Pattern Recognition*, pp. 105–112, Oxford University Press (1995).
10. D. Rimm, R. Camp, L. Charette, J. Costa, D. Olsen, M. Reiss, "Tissue Microarray: A New Technology for Amplification of Tissue Resources", in *Cancer Journal*, Vol. 7, pp. 24–31, 2001.
11. R.L. Camp, L.A. Charette, D.L. Rimm, "Validation of tissue microarray technology in breast carcinoma", in *Laboratory Investigation*, Vol. 80, Issue 12, pp. 1943–1949, Dec. 2000.
12. G.G. Chung, E.P. Kielhorn, D.L. Rimm. 2002. "Subjective differences in outcome are seen as a function of the immunohistochemical method used on a colorectal cancer tissue microarray" in *Clin. Colorectal Cancer*, Vol. 1, pp. 237–242.
13. G.G. Chung, E. Provost, E.P. Kielhorn, L.A. Charette, B.L. Smith, D.L. Rimm, "Tissue microarray analysis of beta-catenin in colorectal cancer shows nuclear phospho-beta-catenin is associated with a better prognosis", in *Clin. Cancer Res.*, Vol. 7, Issue 12, pp. 4013–4020, Dec. 2001.
14. `http://www.cri-inc.com/instruments/products/imaging_varispec.shtml`
15. `http://www.cri-inc.com/`
16. B. Weyn, G. Van der Wouwer, M. Koprowski, A. Van Daele, K. Dhaene, P. Scheunder, W, Jacob, E. Van Marck, "Value of Morphometry, Texture Analysis, Densitometry and Histometry in the Differential Diagnosis of Malignant Mesolthelioma", in *Journal of Pathology*, Vol. 189, Issue 4, pp. 581–589, Dec. 1999.
17. G.L. Mutter, J.P. Baak, C.P. Crum, R.M. Richart, A. Ferenczy, W.C. Faquin, "Endometrial Precancer Diagnosis by Histopathology, Clonal Analysis and COmputerized Morphometry", in *Journal of Pathology*, Vol. 190, Issue, 4, pp. 462–469, Mar. 2000.
18. R. Porter, K. McCabe and N. Bergmann, "An Applications Approach to Evolvable Hardware", in *Proc. of the First NASA/DoD Workshop on Evolvable Hardware*, Pasadena, California, July 19–21, 1999, pp. 170–174.
19. S. Perkins, N.R. Harvey, S.P. Brumby and K. Lacker, "Support Vector Machines for Broad Area Feature Extraction in Remotely Sensed Images", in *Proc. SPIE 4381*, pp. 286–295, 2001.

CANCER
CYTOPATHOLOGY

Diagnostic Classification of Urothelial Cells in Urine Cytology Specimens Using Exclusively Spectral Information

Rajesh Jaganath, B.S.[1]
Cesar Angeletti, M.D., Ph.D.[1]
Richard Levenson, M.D.[2]
David L. Rimm, M.D., Ph.D.[1]

[1] Department of Pathology, Yale University School of Medicine, New Haven, Connecticut.

[2] CRI, Inc., Woburn, Massachusetts.

Supported by a grant from the Patrick and Catherine Weldon Donaghue Foundation and the U.S. Army Breast Cancer Research Program Idea Award DAMD17-02-1-0634 to David Rimm and by the NIH R44 CA88684 Phase II Supplement to Richard Levenson at CRI, Inc.

Address for reprints: David L. Rimm, M.D., Ph.D., Department of Pathology, Yale University School of Medicine, P.O. Box 208023, New Haven, CT 06520-8023; Fax: (203) 737-5089; E-mail: david.rimm@yale.edu

Received November 19, 2003; revision received March 17, 2004; accepted March 24, 2004.

BACKGROUND. Although cytologic evaluation of urine specimens is a standard procedure in the diagnosis and follow-up of bladder carcinoma, its sensitivity and specificity are low. Cytopathologic diagnoses are driven primarily by spatial relations or morphology. Although color enhances the pathologist's perception of the specimen, spectral information plays a minimal role in diagnostic processes. Recently, methods have been developed to capture and analyze spectral information from clinical specimens. In the current study, the authors determined the classification value of spectral information by testing its ability to discriminate between malignant and benign urothelial cells in cytology specimens.

METHODS. Multiple images of benign urothelial cells (n = 39) and urothelial carcinoma cells (n = 35) were collected at serial wavelengths using a liquid crystal tunable optical filter and composited into a mosaic using ENVI (Environment for Visualizing Images) software. Through minimum noise fractionation and principal component analysis, the spectral information in the mosaic was compressed into a 29-dimensional scatter plot. The data generated were analyzed using visual and spectral end member extraction on both the original data set and a second independent data set (test set).

RESULTS. One area of spectral clustering in the scatter plot segmented with carcinoma cells exclusively (100% specific), but was not present in every cell (approximately 50%), which may indicate that these spectral profiles are present in a subpopulation of malignant cells or at specific points of their cell cycle. Using ENVI algorithms, the authors found that a particular classification spectrum (end member 9) and its closest relatives identified malignant cell clusters, with a sensitivity and specificity that reached 82% and 81%, respectively. To validate this mechanism in a test set, a second mosaic comprised of 15 benign and 15 malignant clusters was analyzed using end member 9, resulting in a combined sensitivity and specificity of 73%.

CONCLUSIONS. The results of the current study demonstrate that spectral information, in the complete absence of morphologic or spatial information, allows discrimination of benign and malignant urothelial cells in routine urine cytology specimens. The authors believe that this novel technology, combined with spatial analysis, has the potential to serve as an ancillary test for improved detection of bladder carcinoma. ***Cancer (Cancer Cytopathol)* 2004;102:186–91.**

KEYWORDS: urothelial cells, urine cytology, urothelial carcinoma, bladder carcinoma.

During the last decade, the incidence of bladder and urinary tract carcinoma in the general population has remained stable at 20 per 100,000,[1] representing the fourth most common malignant dis-

DOI 10.1002/cncr.20302
Published online 30 April 2004 in Wiley InterScience (www.interscience.wiley.com).

ease reported among males. The majority of bladder tumors are urothelial carcinomas,[2] which arise from the transitional epithelium at the mucosal surface. Multifocality, a common feature of these neoplasias, is responsible for the high rate of recurrence of bladder carcinoma.[3] Cytologic evaluation of voided urine and bladder washings is one of the standard procedures for the diagnosis and follow-up of this disease.[4,5] However, the sensitivity of these studies is low, ranging between 60% and 80%,[6,7] and the rate of false-negative results is high.[8] Noninvasive or low-grade tumor specimens are the greatest challenge to the cytopathologist, who may be forced to use ambiguous terms such as "atypical urothelial cells," which sometimes leads to improper patient management.[9,10] In the last few years, various new tests for the detection of bladder carcinoma in urine specimens have been commercialized,[11] but to our knowledge none has gained broad acceptance or been definitively shown to exceed the results obtained by conventional cytology.

Cytopathologic diagnoses are driven primarily by spatial relations or morphology. Spectral information (namely color) is relegated to a very minor role as a diagnostic tool. This most likely is because of the inability of the human eye to accurately and reproducibly extract information from actual spectral properties of the stained specimen. The trichromatic theory of color vision[12] postulates that the human retina has receptors for red, green, and blue and that combinations of these signals will recreate any color stimulus.[13,14] For example, pink can be generated either by a mixture of white light and an orange hue, or a mixture of red and cyan, or a mixture of violet, green, and red. Each of these combinations has a different spectroscopic signature that is not detected by the eye. Spectral imaging is a relatively novel technique that is capable of quantitatively measuring spectra in the visible wavelength range from digital images. Several modalities of spectral imaging systems adaptable to biomedical studies have been described in the literature.[15] A popular system is comprised of coupling a liquid crystal tunable optical filter (LCTF) with a regular light microscope.[16] The LCTF uses electrically controlled liquid crystal elements that transmit certain wavelengths while blocking others and it can operate in the visible range (420–720 nanometers [nm]) at intervals as narrow as 1 nm. This technology can be used to construct full spectral traces for each pixel in a captured image. The quantitative information gathered can be used to discover associations between spectral signatures and biologic properties of the imaged cells or tissue specimens. In the current study, we performed spectral analysis of routine urine cytology specimens to determine its ability to discriminate between benign and malignant urothelial cells.

MATERIALS AND METHODS

Specimen Selection

Urine specimens (voided urine and bladder washings) from patients with or without urothelial carcinoma were received between 1996 and 1997. Papanicolaou-stained slides of these specimens were collected from the archives in the Department of Pathology at Yale-New Haven Hospital (New Haven, CT). Up to five representative clusters in each slide were selected by a board-certified cytopathologist.

Data Collection

The collection of 29 images from each selected cluster was performed at 10-nm wavelength intervals (between 420–700 nm) using a BH-2 Olympus light microscope (Olympus America Inc., Melville, NY) equipped with a CRI VariSpec VIS2-CM LCTF (CRI Inc., Woburn MA). The filter wavelength was controlled through a CRI Varispec filter (CRI). The exit of the tunable filter was connected to a Qimaging Retiga 1300 mono CCD camera (Qimaging, Burnaby, British Columbia, Canada) for the collection of processed images. Both the CCD camera and the Varispec were interfaced to a computer and controlled with CRI Pankroma software (CRI) for image acquisition. Using this software, spectral imaging was performed by automatically collecting images at 29 wavelengths from 420 nm to 700 nm at 10-nm intervals. A band-by-band autoexposure routine was used to adjust exposures to compensate for wavelength-sensitive variations in source intensity, filter transmission efficiency, and camera sensitivity. Band-by-band flat-fielding was performed to compensate for unevenness in illumination.

Mosaic Formation

Spectral "stacks" for each cluster imaged were processed using ENVI™ (Environment for Visualizing Images) software (Research Systems, Inc., Boulder, CO) to produce a spectral mosaic containing multiple, separately imaged clusters combined into a single composite spectral stack for the purpose of analysis. ENVI software, used for the remainder of the analyses in the current study, was designed for use in the analysis of satellite images, but can be used for other applications as well. In the current study cell clusters were masked to exclude them from nonepithelial cells that may have been inadvertently included during image capture. These selected clusters were then combined into a single mosaic containing an upper portion consisting of clusters of malignant cells only and a lower

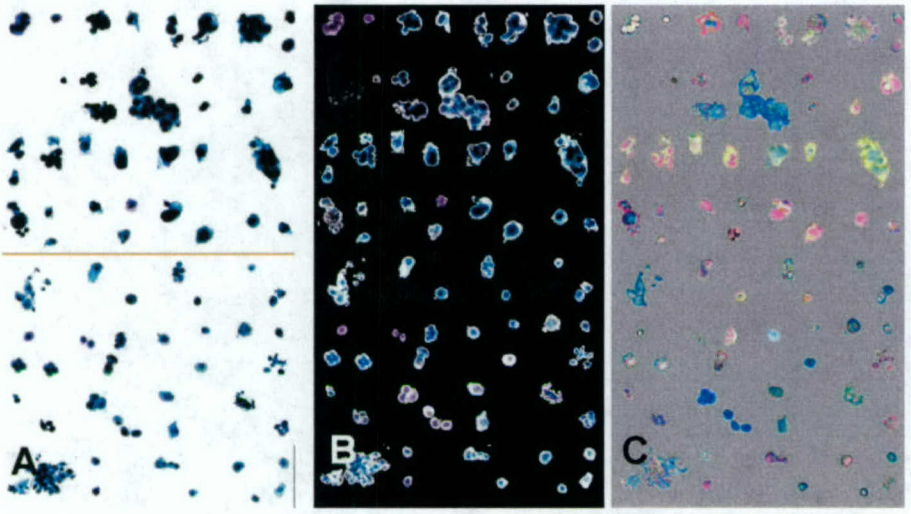

FIGURE 1. Mosaic construction. (A) Initial mosaic containing 29 overlapped spectral images (spectral cube) from 35 clusters of high-grade urothelial carcinoma cells (top half) and 39 clusters of benign urothelial cells. (B) Same mosaic subjected to binary masking to extract background information. (C) Minimum noise fractionation and principal component analysis.

portion with clusters obtained from benign urine specimens. A common white background was generated through data collection of an unstained ThinPrep® (Cytyc Corporation, Boxborough, MA) slide.

ENVI Transforms and *n*-Dimensional Scatter Plot Generation

A binary mask based on spectral characteristics was then applied to the mosaic. The spectral mask assigned a value of "1" to all pixels with intensities between 0 (black) and 220 (of 255) and a value of "0" to the remainder of the mosaic, corresponding to the white background. The masked mosaic was then subjected to a minimum noise fraction (MNF) transform to maximize the variance within the sample set as part of the segmentation process attempting to separate ("segment") the malignant from benign clusters. The MNF transform included in the ENVI software is a sequential two-part principal components transformation. The first transformation, based on an estimated noise covariance matrix, decorrelates and rescales the noise in the data so that the noise has unit variance and so that there is no correlation between bands. Only the bands with data are used for further analysis. This was followed by a principal components transformation, which generates uncorrelated output bands, isolates noise components, and decreases the dimensionality of the data sets. Uncorrelated output bands are found by using a new set of orthogonal axes that have their origin at the data mean but are rotated to maximize data variance. The resulting data were visualized as an *n*-dimensional scatter plot, with each coordinate representing a function of reflectance values in each band for a given pixel.

n-Dimensional Scatter Plot Analysis

The MNF transformed data were subjected to the ENVI pixel purity index (PPI). This process projects the data cloud along random vectors multiple times (5000), records the values at each iteration, and selects the most spectrally pure values. By means of the ENVI spectral angle mapping (SAM) function, all the vectors in the scatter plot were compared with each end member to determine their spectral angle, which is an estimation of the similarity between them. These similarities can be visualized in the mosaic as a blue–pink pseudogradient. The mosaic image was then exported to Adobe Photoshop (Adobe, Mountain View, CA) and forced into a three-color (magenta, cyan, black) indexed image. The number of magenta pixels in each cluster, corresponding to the spectral profiles more closely related to the end member, was calculated as the percentage of the total number of pixels per cell cluster.

RESULTS

Figure 1A shows an image of the final mosaic containing images of cellular clusters captured from 44 urine cytology specimens. Thirty-five clusters of high-grade urothelial carcinoma cells were positioned at the top half of the mosaic and 39 cellular clusters from benign urine specimens were located on the bottom half. First, we extracted the background information through binary masking (Fig. 1B) and subjected the remaining image to MNF and principal component analysis (PCA) (Fig. 1C). A two-dimensional projection of a higher dimensional scatter plot obtained through PCA of all the spectral profiles in the mosaic is depicted in Figure 2A. ENVI software allows multiple

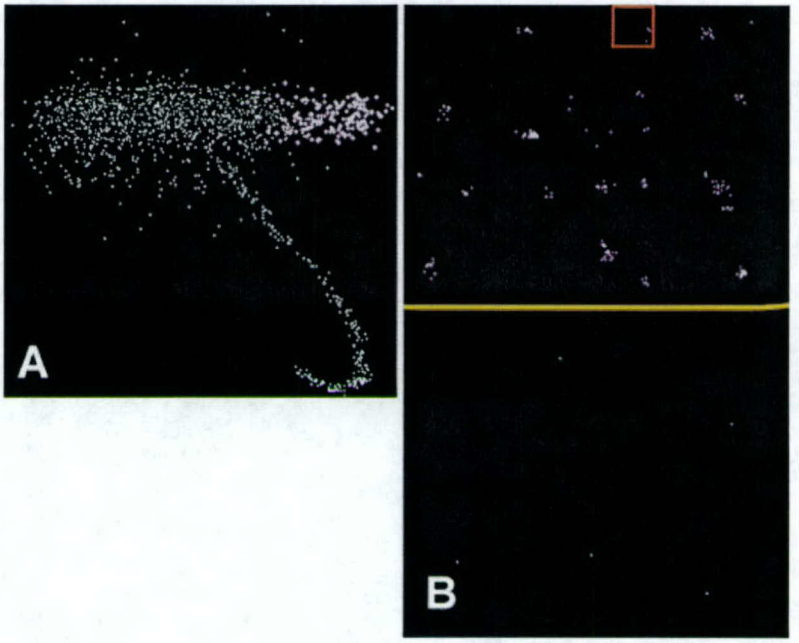

FIGURE 2. (A) *n*-Dimensional scatter plot. Each data point represents a unique spectral profile. The highlighted pink spots in the upper right panel corresponds to one area of spectral clustering. (B) Distribution of the highlighted vectors in the scatter plot when exported back to the mosaic (see fig. 1A) to estimate their distribution among the cellular clusters.

serial bidimensional views of the scatter plot (which, in fact is 29-dimensional). The careful observation of these views revealed the presence of a few areas with higher data point density, such as the one highlighted in Figure 2A. These areas of spectral clustering were exported back to the mosaic to evaluate their distribution among the cells. As a general rule, a cell cluster was considered "identified" if it contained three or more pixels with spectral profiles from these areas. For instance, the area highlighted in Figure 2A identified malignant cell clusters exclusively, but not all of them (Fig. 2B). This particular region in the scatter plot identified 18 of the 35 malignant cell clusters (51.4%).

As an alternative approach, the MNF-transformed data were subjected to PPI analysis, resulting in 30 end members, which correspond to the most spectrally pure pixels (Fig. 3A). These end members were used as references to which the vectors from other pixels in the mosaic could be compared using the ENVI SAM function. From these 30 end members, *n*-dimensional class 9 (indicated with an arrow in Fig. 3B) resulted in the highest segmentation between the nonneoplastic and urothelial carcinoma cell clusters. Figure 3B shows the spectral relation of every pixel in the mosaic to end member 9 and its distribution in the mosaic. The pixels with more spectral similarity to this end member are highlighted in grades of pink and appear to be preferentially located in the top half, corresponding to the malignant cells. The mosaic was converted to indexed color mode in Adobe Photoshop by which every pixel in the image can be forced into the color categories cyan and magenta. Based on the proportion of magenta pixels per cluster, we compared the sensitivity and specificity of this system to identify malignant urothelial cell clusters, creating a receiver operator characteristic (ROC) curve (Fig. 4). We found that the highest combined sensitivity and specificity reached 82% and 81%, respectively, if we assumed that a cluster was malignant when $\geq$ 60% of its pixels were magenta (Fig. 4A).

To demonstrate that this spectral classification was not specific to a trivial feature of the small set we chose for initial analysis, a second mosaic was created with a new set of 15 benign and 15 malignant cell clusters obtained from 9 urine cytology specimens not used in the training set (3 benign and 6 malignant specimens). This smaller "test set" then was analyzed using end member 9 as derived from the first mosaic. The resulting ROC curve is depicted in Figure 4B. Using the same rule of 60% magenta pixels to classify malignant cells, the highest combined sensitivity and specificity was 73%, suggesting that this end member was identifying spectral features associated with malignancy rather than some other more trivial property of the set of cells in the training set.

DISCUSSION

Although spectral imaging technology was designed initially for the analysis of satellite images, we have found that the spectroscopic analysis of routinely stained cytology slides can be highly informative. The results in the current study suggest that spectral data alone are capable of classifying urothelial cells as benign or malignant.

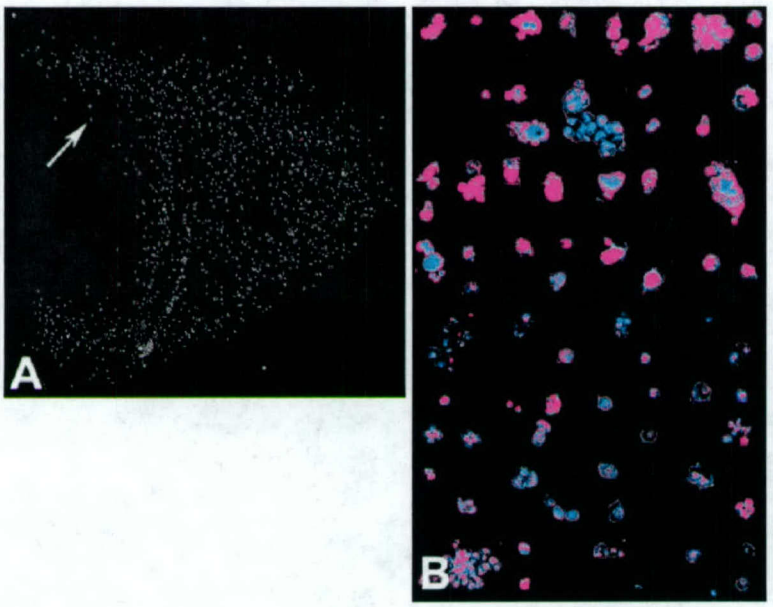

FIGURE 3. (A) Scatter plot subjected to the purity pixel index (PPI). End-members are highlighted in different colors. End-member #9 (magenta) is indicated by an arrow. B) Spectral angle mapping. Similarities between every pixel in the mosaic and end-member #9 are depicted as a pseudo-gradient of pink (smaller spectral angle) and blue (higher spectral angle) colors (not shown). Then the mosaic was forced to a three-color indexed image (magenta, blue and black) using Adobe Photoshop for quantification.

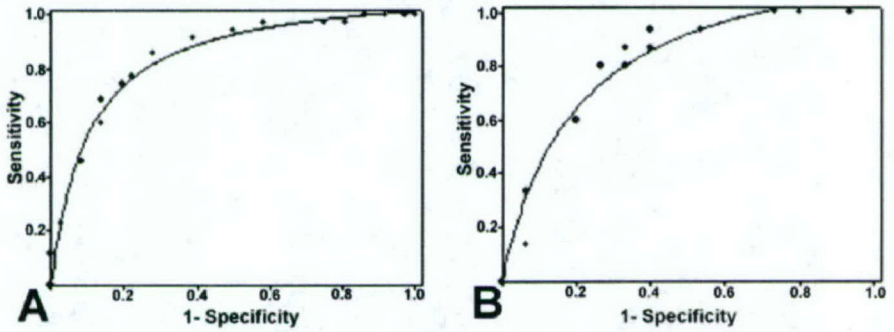

FIGURE 4. (A) Receiver operator characteristic (ROC) curve of the first mosaic with end member 9 as the reference data point. For the purpose of generating the curve, a cluster was classified as malignant when 5–100 % (at 5% intervals) of its pixels were closely related to end member 9. (B) ROC curve of a "test" mosaic using similar rules for end member 9.

Because "obviously" cytologically benign or malignant cells were used for the training and test sets, the spectral method is not yet equivalent to conventional spatial/morphologic analysis. Nonetheless, we are encouraged by the amount of information obtainable in the complete absence of information regarding spatial relations of any given feature to another (such as nuclear size, shape, nuclear-to-cytoplasmic ratio, and clustering).

Spectral analysis may be subject to a whole new set of preparative variables. When a mixture of dyes is used, such as the Papanicolaou stain, the differential binding to cellular components results in unique spectral profiles, depending on the relative amounts of individual dyes present and the shifts in spectral properties.[17] However, in our tests of Papanicolaou-stained urine specimens collected from the routine workload cases from different days and different batches of stain, the staining process appears to be sufficiently homogeneous for spectral analysis. Therefore, the molecular interactions that generate the spectral signatures appear to be robust with respect to minor batch-to-batch variations. We have not yet tested specimens from other laboratories. Furthermore, the Papanicolaou stain appears to be a spectrally rich choice. An unpublished analysis of a series of eight histologic stains performed by our laboratory suggested that Papanicolaou, hematoxylin and eosin, and Feulgen stains were the promising candidates for use in spectral imaging.

The optimal method for analysis of spectral data for cytopathology remains to be determined. Visual

analysis of the scatter plot revealed a few areas with higher spectral profile density. One of these areas (Fig. 2A) contained vectors that were only present in malignant cells. However, approximately one-half of the malignant cells tested did not express these spectral profiles, suggesting that they may be present in just a subpopulation or subclone of malignant cells. Oligoclonality in urothelial carcinoma has been reported previously, and may be related to the high tendency of these tumors to synchronous and metachronous multifocality.[18,19] Another plausible explanation is that these unique color properties are common to all urothelial carcinoma cells, but at specific points of their cell cycle. In agreement with that, Holman et al.,[20] using infrared spectroscopy on a cell culture system, reported changes in the spectral absorption of cells at different phases of the cell cycle. They proposed that these changes were related to variations in the DNA/RNA content and protein structure, which could also affect the distribution of dye mixtures.

A second approach to the analysis of the scatter plot involved extraction of end members or "spectrally pure" pixels. The ENVI software can localize end members by repeatedly projecting *n*-dimensional scatter plots to random unit vectors and recording the number of times that each pixel is marked as extreme. Purest end members will tend to land at or near the boundaries of an *n*-dimensional scatter plot, and thus will have the highest "extreme" score. Using SAM, we found that pixels with the highest similarity to end member 9 segmented preferentially in the clusters of malignant cells. By applying different rules of classification based on the relative number of pixels per cell cluster, we could generate an ROC curve with a combined sensitivity and specificity of $\leq$ 81% and 82%, respectively (Fig. 4A) for the training set and a combined sensitivity and specificity of $\leq$ 73% (Fig. 4B) for the test set. We do not know whether this method represents the optimal way to analyze these data, but it represents the most rigorous method we could obtain for this type of data set. We currently are developing new methods that would be more specific for data obtained from histologic rather than satellite images and, most promisingly, we are working toward the incorporation of spatial and spectral information.

The results of the current study indicate that spectral information alone is capable of classifying malignant and benign urothelial cells in urinary cytology specimens. We believe that this novel technology, most likely combined with spatial analysis, has a great potential to become a powerful ancillary or primary diagnostic test for cytopathology specimens.

REFERENCES

1. http://seer.cancer.gov/csr/1975_2000/results_merged/sect_26_urinary_bladder. pdf [Accessed July 28, 2003].
2. Johansson SL, Cohen SM. Epidemiology and etiology of bladder cancer. *Semin Surg Oncol.* 1997;13:291–298.
3. Rodriguez-Alonso A, Pita-Fernandez S, Gonzalez-Carrero J, Nogueira-March JL. Multivariate analysis of survival, recurrence, progression and development of mestastasis in T1 and T2a transitional cell bladder carcinoma.. *Cancer.* 2002; 94:1677–1684.
4. Messing E. Urothelial tumors of the urinary tract. In: Walsh PC, Vaughan ED, et al., editors. Campbell's urology. St. Louis: W.B. Saunders, 2002:2732–2765.
5. Tut VM, Hildreth AJ, Kumar M, Mellon JK. Does voided urine cytology have biological significance? *Br J Urol.* 1998; 82:655–659.
6. Gregoire M, Fradet Y, Meyer F, et al. Diagnostic accuracy of urinary cytology and deoxyribonucleic acid flow cytometry and cytology on bladder washings during followup for bladder tumors. *J Urol.* 1997;157:1660–1664.
7. Kumar NU, Dey P, Mondal AK, Singh SK, Vohra H. DNA flow cytometry and bladder irrigation cytology in detection of bladder carcinoma. *Diagn Cytopathol.* 2001;24:153–156.
8. Bastacky S, Ibrahim S, Wilczynski SP, Murphy WM. The accuracy of urinary cytology in daily practice. *Cancer.* 1999; 87:118–128.
9. Curry JL, Wojcik EM. The effects of the current World Health Organization/International Society of Urologic Pathologists bladder neoplasm classification system on urine cytology results. *Cancer.* 2002;96:140–145.
10. Renshaw AA. Subclassifying atypical urinary cytology specimens. *Cancer.* 2000;90:222–229.
11. Lotan Y, Roehrborn CG. Sensitivity and specificity of commonly available bladder tumor markers versus cytology: results of a comprehensive literature review and meta-analyses. *Urology.* 2003;61:109–118.
12. von Helmoltz H. In: Southhall JPC, editor. Treatise on physiological optics. New York: Dover Publications, 1962.
13. Lennie P. Color vision: putting it together. *Curr Biol.* 2000; 10:R589–591.
14. Roorda A, Williams DR. The arrangement of the three cone classes in the living human eye [comment]. *Nature.* 1999; 397:520–522.
15. Farkas DL, Du C, Fisher GW, et al. Non-invasive image acquisition and advanced processing in optical bioimaging. *Comput Med Imaging Graph.* 1998;22:89–102.
16. Ornberg RL, Woerner BM, Edwards DA. Analysis of stained objects in histological sections by spectral imaging and differential absorption. *J Histochem Cytochem.* 1999;47:1307–1314.
17. Rothmann C, Barshack I, Kopolovic J, Malik Z. Spectrally resolved morphometry of the nucleus in hepatocytes stained by four histological methods. *Histochem J.* 1998;30: 539–547.
18. Cheng L, Gu J, Ulbright TM, et al. Precise microdissection of human bladder carcinomas reveals divergent tumor subclones in the same tumor. *Cancer.* 2002;94:104–110.
19. Sidransky D, Frost P, Von Eschenbach A, et al. Clonal origin bladder cancer. *N Engl J Med.* 1992;326:737–740.
20. Holman HY, Martin MC, Blakely EA, Bjornstad K, McKinney WR. IR spectroscopic characteristics of cell cycle and cell death probed by synchrotron radiation based Fourier transform IR spectromicroscopy. *Biopolymers.* 2000;57:329–335.

Spectral-Spatial Analysis of Urine Cytology Specimens

Authors:
Cesar Angeletti[1], Neal Harvey[2], Vitali Khomitch[1], Andrew Fischer[3], Richard Levenson[4] & David L. Rimm[1]

Addresses:

1) Department of Pathology, Yale University School of Medicine, P.O. Box 208023 New Haven, Connecticut 06520

2) Space and Remote Sensing Sciences Group, International, Space and Response Technologies Division, Los Alamos National Laboratory, Mail Stop D-436, Los Alamos New Mexico 87544

3) Department of Pathology
University of Massachusetts Medical School, 365 Plantation Street, Worcester, MA 01605

4) CRI, Inc., 35-B Cabot Rd., Woburn, Massachusetts 01801

Correspondence should be addressed to:
David L. Rimm, M.D., Ph.D.
Department of Pathology
Yale University School of Medicine
PO Box 20803
310 Cedar St.
New Haven, CT 06520
Phone: (203) 737-4204
Email: david.rimm@yale.edu

ABSTRACT

Despite low sensitivity (around 60%) cyto-morphologic examination of urine specimens represents the standard procedure in the diagnosis and follow-up of bladder cancer. Although color is information-rich, morphologic diagnoses are rendered almost exclusively on the basis of spatial information. We hypothesized that quantitative assessment of color (more precisely, of spectral properties) using liquid crystal-based spectral fractionation, combined with genetic algorithm-based spatial analysis, can improve on the accuracy of traditional cytologic examination. Images of various cytological specimens were collected every 10 nm from 400 to 700 nm to create an image stack. The resulting datasets were analyzed using the Los Alamos-developed GENIE package (GENetic Imagery Exploitation, a hybrid genetic algorithm that segments (classifies) images using automatically "learned" spatio-spectral features. In an evolutionary fashion, GENIE generates a series of algorithms or "chromosomes", keeping the one with best fitness with respect to a userdefined training set. When targeted to detect malignant urothelial cells in cytology specimens, GENIE showed a combined sensitivity and specificity of 85% and 95%, in samples drawn from two institutions over a span of 4 years. When trained on cells from biopsy-adjudicated specimens initially diagnosed as "atypical," GENIE showed efficiency superior to the cytopathologist with respect to predicting the biopsy result. In the future, this method could be used as an ancillary test in cytopathology, used in a manner analogous to immunostaining, but without any further processing, since the information is derived from samples stained with the routine Papanicolaou stain. It could help to provide a more accurate diagnosis in those situations when a definitive diagnosis cannot be rendered based solely on the cyto-morphology.

INTRODUCTION

Urothelial carcinoma is diagnosed and followed by cytological evaluation of voided urine and bladder washings[1,2]. The sensitivity of these tests is low, ranging between 60 and 80%.[3,4]. Non-invasive or low-grade tumors offer an even greater challenge to the cytopathologist[5], resulting in the use of such terminology as "atypical urothelial cells," leading to suboptimal patient management[6].

Morphologic diagnoses in cytopathology are primarily driven by spatial relationships, while color-based information is relegated to a very minor role. This may be because the human eye has limited spectral resolution, viewing the world in broad, overlapping spectral bands of red, green and blue. If the spectral properties of cytological specimens are important, this fact can only be assessed using devices capable of imaging with spectral precision. This excludes conventional red-green-blue (RGB) cameras, which have no more spectral resolving power than the human visual system they are designed to emulate. Spectral imaging, on the other hand, is a relatively novel technique capable of quantitatively measuring optical spectra on a pixel-by-pixel basis, and therefore capturing differences normally overlooked by the human eye[7]. Previous studies have shown that spectral analysis has high information content, albeit less than spatial information[8]. Thus, combination of spectral imaging with spatial analysis (morphology) could maximize information obtainable from routine pathology slides and could serve as a valuable ancillary test in diagnostic pathology.

GENetic Imagery Exploitation or "GENIE," is a recently developed artificial intelligence platform that allows optimization of images analysis algorithms through selection of learned spatio-spectral features[9]. This system was originally conceived at Los Alamos National Laboratory for the analysis of satellite images of the surface of the earth[10]. GENIE is a hybrid learning system that combines a genetic algorithm that searches in a space of image-processing operations for a set that can produce suitable feature planes, coupled to a more conventional classifier that uses those features planes to output a final classification. In an evolutionary fashion, GENIE generates a series of algorithms or "chromosomes", and assigns an individual fitness score related to how well they classify a set of training images. Each chromosome is composed of a variable number of mathematical operators or "genes" that are interpreted serially. At the end of each generation, GENIE selects chromosomes to participate in processes of crossover and mutation, with the probability of selection being based on their fitness, and the cycle is repeated. When compared to other supervised classifiers for multispectral image feature analysis, GENIE out-performed them all in almost every task tested, indicating higher sensitivity and generalization abilities[11].

Here, we first applied the GENIE hybrid genetic algorithm to multispectral images obtained from artificially produced Papanicolaou-stained cytology slides, to test its capabilities to detect malignant cells. Then, we extended our study to routine urine cytology slides from different time-periods and separate institutions. Finally, we tested its performance on biopsy-adjudicated clinical specimens initially called "atypical" by the cytopathologist, to test its potential value as an ancillary test to increase the accuracy of cytopathology.

RESULTS

Detection of malignant cells in defined-composition cytology specimens

The first analysis was done to determine if this technology could define cancer cells from normal cells, in a system free of pathologist interpretation. Colon adenocarcinoma and benign colonic epithelial cells were scraped from two separate areas of a colectomy specimen, and

mixed in preservative at various concentrations. From each mixture, one ThinPrep™ slide was prepared, and all slides were stained together using the Papanicolaou stain. GENIE training was performed using 5 high-power (40X) multispectral images (420 to 700 nm, 10 nm intervals) of the slides prepared from the vials containing 0% and 100% tumor cells. Training data was created by manually marking the cells in images in red (for "non-feature" or benign) and green (for "feature" or malignant). 300 generations were run under specific training conditions (Supplementary Note 1). A chromosome was generated (# 004345), with a fitness with respect to the training set of 835 (a score of 1000 would be consistent with every pixel in the training data correctly classified). Twenty multispectral test images acquired from every slide were run in GENIE using chromosome #004345. The results showed most of the pixels corresponding to tumor cells clearly identified as "feature" (green color overlaying the original images), while those corresponding to benign colonic epithelial cells were ignored (red color, see Fig.1**a**). In a separate training session, we created another chromosome that identifies background areas by manually indicating areas of background as "feature," while any cell (benign and malignant) was indicated as "non-feature." We combined the results of both chromosomes so as to mask the background, and exclusively score the results on cell-associated pixels. This was necessary because: 1) chromosome #04345 was trained with spectro-spatial information from cell-associated areas only, so it should be evaluated on that same substrate; 2) the surface occupied by cells was variable between images. The scoring was done by calculating the proportion of pixels recognized as "tumor" in the total cell surface of the images (GENIE Index). The average GENIE index obtained for each ThinPrep™ slide increased linearly according to the percentage of tumor cells in the original suspensions, with a correlation coefficient of 0.84 (Fig.1**b**). Note that the adhesive quality of the cells and subsequent clustering made it impossible to get homogeneous dispersions on the slides, partially limiting the ability to evaluate this assay. The accuracy of this assay could have been improved by using more training images, but having proved the concept, we moved on to actual cytology specimens.

Detection of malignant urothelial cells in routine urine cytology specimens

To create a chromosome able to detect malignant urothelial cells in routine urine specimens, we trained GENIE with 12 Papanicolaou stained ThinPrep™ urine cytology slides received in the Department of Pathology at Yale-New Haven Hospital between 1996 and 1997. From these slides, 38 high-power multispectral images were acquired and imported into the GENIE platform. 161 nuclei from benign urothelial cells and 70 nuclei from malignant urothelial cells were selected by a cytopathologist as "non-feature" and "feature," respectively. Cells showing atypia, but lacking definitive features of malignancy, as well as inflammatory cells and degenerated urothelial cells, were not selected. The training conditions used are listed in Supplementary Note 2. GENIE produced a single best chromosome, #025867, containing four genes, which was evaluated on two validation sets. The first set was constructed from 17 unique patients' slides with a diagnosis of urothelial carcinoma received in the Department of Pathology at Yale New Haven Hospital between 1998 and 1999. The set contained a total of 190 benign urothelial cells and 178 malignant urothelial cells. GENIE identified malignant cells by overlaying green color over their nuclear areas (Fig. 2 **center column**). Using 90% of the nuclear area being identified as "feature" as the cut-off value, this chromosome classified malignant and benign urothelial cells with a sensitivity and specificity of 87% and 96%. The selection of 90% as the cut off value for nuclear area was arbitrary and could be optimized for even more accurate scoring in the future. It was selected here to be a conservative example for illustration. Atypical or degenerate cells were not used in the calculation, since they were excluded from the training

set. Inflammatory cells were also excluded; however this chromosome did not identify neutrophils as "feature."

The second validation set was composed of eight cases (6 positive for urothelial carcinoma and two negative for malignancy) from the Department of Pathology at University of Massachusetts Medical School. The Papanicolaou staining procedure routinely used at both institutions is similar (see Methods section). This test set comprised 121 benign urothelial cells and 40 malignant urothelial cells. Chromosome #025867 demonstrated a sensitivity and specificity of 85% and 96%.

Finally, to test the importance of the spectral information in this chromosome we altered it in a fashion analogous to a transposition mutation where information from two spectral data planes was exchanged. Specifically, the mathematical operator that used a 550 nm plane was given the 570 nm plane and vice versa. Under these new conditions, Chromosome #025867 misclassified several benign nuclei as malignant (Fig 2, **right column**), further illustrating the critical nature of the spectral component of the analysis.

Detection of malignant urothelial cells in "atypical" urine cytology specimens

Adjudication of cases defined as "atypical" represents a critical challenge for pathologists and a potential clinical application of this technology. To investigate this possibility further, we selected sixteen unique urine cytology cases initially diagnosed as "atypical" by a cytopathologist and received at Yale-New Haven Hospital between 1995 and 1996. Eight slides (negative follow-up group) corresponded to cases that had at least one specimen subsequent to the index case where the cytology, biopsy or surgical specimen was diagnosed as "negative for malignancy" in the following year. Cases were excluded if subsequent specimens included "atypical", "suspicious" or "positive for malignant cells" in the diagnostic text. The other eight (positive follow-up group) had least a one specimen subsequent to the index case where the biopsy or surgical specimen was called "positive for malignant cells" in the following year. A single image containing the most atypical cell cluster was used for each case. Areas corresponding to the atypical clusters of the "follow-up positive group" were manually designated as "feature", while the ones of "follow- up negative group" as "non-feature." The training conditions used were similar to those used for chromosome #025867 (see above and supplemental materials). The fitness of this chromosome with respect to the training set was 824. We then prepared a completely unique validation set using images from 34 "negative follow-up" and 51 "positive follow-up" urine cytology cases from archival material received between 1997 and 2002. The criteria used in the selection of cases were similar to those for the training set. One atypical cell cluster per slide was selected for acquisition and each sample corresponded to a different patient. Scoring was done by calculation of the GENIE index in each cell cluster alone. The results are shown as a frequency distribution of GENIE index scores (Figure 3**a-b**). The positive follow-up group shows a significantly higher mean score than the negative follow-up. There is no cut-point that definitively separated the groups, but an ROC (Receiver Operator Characteristic) curve could be generated based on the number of pixels per cell cluster used to define a positive identification. The area under the ROC curve (AUC) obtained was 0.728 (Fig **4**).

A problem with this analysis is that we could be using the wrong cells to train GENIE. That is, although we know the biopsy outcome of each case, we do not know if the clusters used for training were actually malignant cells. Since single clusters could be mis-assigned, we attempted to improve the performance of GENIE on atypical urine cytology cases by re-training the system using a larger number of training images (32 cases, including those used in the

previous training session). The resulting chromosome, #020105, had a fitness with respect to the training set of 732. When this chromosome was run on the test set, it performed better compared to # 026897 (Fig 3**c-d**), with an area under the ROC curve of 0.784 (Fig **4**).

DISCUSSION

GENetic Imagery Exploitation, or "GENIE," is a newly developed artificial intelligence system that performs spatio-spectral analysis of images. This is in contrast to a long history of quantitative analysis systems tested on urine specimens which all have used spatial information, and monochromatic (or panchromatic) optical density[12-15]. While some showed promising results, none have been widely adopted due to their inability to definitively improve on the current, manual, standard. GENIE uses spectral data to increase the total information content. Our previous work (ref) and the "mutagenesis" experiments above illustrate the value of this additional information. Note that the colors selected by GENIE are from a range of 18 wavelengths and are not those that would create a standard RGB (red-green-blue) image. The fact that certain wavelengths are more informative than others suggests subtleties in protein-dye interactions that are impossible to appreciate with the human eye.

A critical difficulty in the work is to compare the objective results of the GENIE analysis to the subjective results of the pathologist. When confronted with urine cytology cases classified as "atypical" by the cytopathologist, GENIE produced a chromosome with a combined sensitivity and specificity of approximately 70% (area under the ROC curve of 0.72) when tested on a cohort of 85 patients. It is difficult to compare the AUC since sensitivity and specificity cannot be calculated from a diagnosis of "atypical". An AUC could be calculated for the pathologist by repeated analysis with forced selection of a positive or negative diagnosis, but this was not done since our goal was to compare GENIE to a "real life" situation. Moreover, while the pathologist's report was generated looking at the whole slide, the GENIE "diagnosis" was based on one single cluster per slide. By doubling the number of training images in a new training session, we generated a chromosome that showed higher sensitivity and specificity on the same test set (area under the ROC curve of 0.78). This suggests that future efforts that include more cells or even the whole slide in training sets may substantially improve the AUC even in this difficult class of cases.

A further limitation of this initial analysis is our using of a single chromosome to identify only positive events. Future studies are underway to use multiple chromosomes where some are trained on other subclasses of cells that can be used to eliminate false positive identifications. For example GENIE could be trained on benign parabasal or umbrella cells and then these chromosomes could be combined in a subtractive manner to determine an ultimate score. Combining algorithms trained to identify more restricted features, and integrating them through a higher order "cognizant" chromosome (incorporating abstract thinking or data fusion) represents an area of intensive research in modern computational analysis of remote sensing data[16]. We have begun applying this concept on cytology samples in new studies in order to evaluate its clinical potential.

METHODS

Colon adenocarcinoma cytology specimen preparation: Colon adenocarcinoma cells and normal colonic epithelial cells were manually scraped from the surface of a fresh colon resection specimen received at the Department of Pathology at Yale. The scraped material was immediately suspended in RPMI medium and gently dispersed using a glass Dounce

homogenizer. Cell count was performed using a hemocytometer. Aliquots from each suspension were mixed in 11 ThinPrep™ vials containing 20 ml of PreservCyt® preservative at different malignant/benign cell ratios, and total cell concentration of $1x10^5$/ml. A ThinPrep™ slide was prepared from each vial in a 2000 ThinPrep™ processor (Cytyc, Boxborough, Massachusetts), and stained with Papanicolaou stain. Random areas containing cells or cell clusters were selected for image acquisition.

Urine cytology specimen selection: Papanicolaou-stained urine cytology ThinPrep™ slides were collected from archival material at the Departments of Pathology of Yale-New Haven Hospital and University of Massachusetts Medical School. Representative areas containing benign, atypical or malignant urothelial cells were selected by a cytopathologist for image data acquisition.

Data collection: Multispectral images were acquired from 420 nm to 700 nm at 10-nm intervals from each selected area at 40X-magnification using a a BH-2 Olympus light microscope (Olympus America, Melville, New York) equipped with a CRI VariSpec™ VIS2-CM liquid crystal tunable filter (CRI, Woburn MA). The tunable filter was coupled to a Retiga 1300 monochrome CCD camera (Quantitative Imaging, Burnaby, British Columbia). Both CCD camera and VariSpec were controlled by CRI acquisition software (CRI, Woburn, Massachusetts). Exposure time for each slide and wavelength was calculated prior to each acquisition. Background information was subtracted by acquiring an area with no cells (one per slide) and flat-fielding over the area of interest. Digital images were produced as arrays of 896 by 768 pixels at 8-bit resolution.

GENIE training and "chromosome" generation: Training image data was introduced into the GENIE analysis by means of the ALADDIN Java-based graphic tool. ALADDIN allows the analyst to select areas to be considered in the training session as "feature" and "non-feature" by manually painting those areas as a colored overlay. The learning system parameters can be modified in terms of number of chromosomes per generation, maximum number of genes in each algorithm, number of generation cycles, mutation rate, recombination rate, and end-point fitness goal.

Detection of malignant cells using GENIE "chromosomes:" Multispectral image data from the test sets was subjected to mathematical transformation by means of algorithms previously generated in the training process. The modified result images could be viewed as green (feature) and red (non-feature) areas overlying the original images (used for orientation purposes), or as black-white binary images (used for scoring purposes). Scoring of the GENIE result images was done by calculating the GENIE Index, which represents the proportion of pixels inside the object of study (nuclei, cell, cluster, etc) recognized as "feature." ROC curves were constructed on Excel_Analyze-it® software (Leeds, England). Combined sensitivity and specificity were calculated using increasing levels of GENIE Index as cut off values.

ACKNOWLEDGMENTS:

This work is supported by grants from the Patrick and Catherine Weldon Donaghue Foundation and US Army Breast Cancer Research Program Idea Award DAMD17-02-1-0634 to David Rimm and by the NIH R44 CA88684 Phase II Supplement to Richard Levenson at CRI, Inc. and by the American Society of Cytopathology Research Seed Award to Cesar Angeletti.

REFERENCES

1. Messing, E. Urothelial tumors of the urinary tract. in *Campbell's Urology* (ed. Walsh PC, R.A., Vaughan ED, Wein AJ, Kavoussi LR, Novick AC, Partin AW, Peters CA) 2732-2765 (WB Saunders, St Louis MO, 2002).
2. Tut, V.M., Hildreth, A.J., Kumar, M. & Mellon, J.K. Does voided urine cytology have biological significance? *British Journal of Urology* **82**, 655-9 (1998).
3. Gregoire, M. et al. Diagnostic accuracy of urinary cytology, and deoxyribonucleic acid flow cytometry and cytology on bladder washings during followup for bladder tumors. *Journal of Urology* **157**, 1660-4 (1997).
4. Kumar, N.U., Dey, P., Mondal, A.K., Singh, S.K. & Vohra, H. DNA flow cytometry and bladder irrigation cytology in detection of bladder carcinoma. *Diagnostic Cytopathology* **24**, 153-6 (2001).
5. Curry, J.L. & Wojcik, E.M. The effects of the current World Health Organization/International Society of Urologic Pathologists bladder neoplasm classification system on urine cytology results. *Cancer* **96**, 140-5 (2002).
6. Renshaw, A.A. Subclassifying atypical urinary cytology specimens. *Cancer* **90**, 222-9 (2000).
7. Farkas, D.L. et al. Non-invasive image acquisition and advanced processing in optical bioimaging. *Comp. Med. Imag. Graph.* **22**, 89-102 (1998).
8. Jaganath, R., Angeletti, C., Levenson, R. & Rimm, D.L. Diagnostic classification of urothelial cells in urine cytology specimens using exclusively spectral information. *Cancer* **102**, 186-91 (2004).
9. Perkins S., T.J., Brumby S.P., Harvey N.R., Porter R., Szymanski J.J., Bloch J.J. GENIE: A hybrid genetic algorithm for feature classification in multi-spectral images. *Proc. SPIE* **4120**, 52-62 (2000).
10. Brumby S.P., T.J., Perkins S.J., Harvey N., Szymanski J.J., Bloch J.J., Mitchell M. investigation of image feature extraction by a genetic algorithm. *Proc. SPIE* **3812**, 24-31 (1999).
11. Harvey N.R., T.J., Brumby S.P., Perkins S., Szymanski J.J., Bloch J.J., Porter R.B., Galassi M., Cody Young A. Comparison of GENIE and conventional supervised classifiers for multispectral image feature extraction. *IEEE Trans. Geosci. Remote Sensing* **40**, 393-404 (2002).
12. Koss, L.G., Bartels, P.H. & Wied, G.L. Computer-based diagnostic analysis of cells in the urinary sediment. *Journal of Urology* **123**, 846-9 (1980).
13. Sherman, A.B. et al. Bladder cancer diagnosis by computer image analysis of cells in the sediment of voided urine using a video scanning system. *Analytical & Quantitative Cytology & Histology* **8**, 177-86 (1986).
14. Van der Poel, H.G. et al. Bladder wash cytology, quantitative cytology, and the qualitative BTA test in patients with superficial bladder cancer. *Urology* **51**, 44-50 (1998).
15. de la Roza, G.L. et al. DNA image analysis of urinary cytology: prediction of recurrent transitional cell carcinoma. *Modern Pathology* **9**, 571-8 (1996).
16. Harvey N.R., P.S.J., Pope P.A., Theiler J., David N.A., Porter R.B. Investigation of automated feature extraction using multiple data sources. *Proc. SPIE* **5099**, 190-200 (2003).

Legends to Figures

Figure 1. Detection of malignant cells in cytology specimens of defined composition using GENIE. a) GENIE results are displayed on the right as binary green (feature) and red (non-feature) images, superimposed to the original image (better appreciated on the left). **A-D**. Testing images taken from the slide containing no colon adenocarcinoma cells. **E-F**. Testing images from the 100 % colon adenocarcinoma slide. b) Quantification of malignant colonic epithelial cells in mixtures of defined composition. Every data point is the average of 20 high-power field (400X) from each slide, prepared from vials containing a defined proportion of malignant cells.

Figure 2. Detection of malignant urothelial cells in routine urine cytology slides using GENIE. Original images are displayed on the left side of the figure. The center column corresponds to GENIE result images using chromosome #04345. This chromosome identifies nuclei of urothelial carcinoma cells by overlaying green color over its surface. The right column shows the performance of chromosome #04345 when spectral data planes D13 and D15 are switched (in an analogous fashion to transposition mutation).

Figure 3. GENIE detection of malignancy in routine urine cytology slides previously diagnosed as "atypical" by a cytopathologist. **A**. Results obtained with chromosome #026897. **B**. Results obtained with chromosome #020105. **C**. ROC curve for both chromosomes on test set.

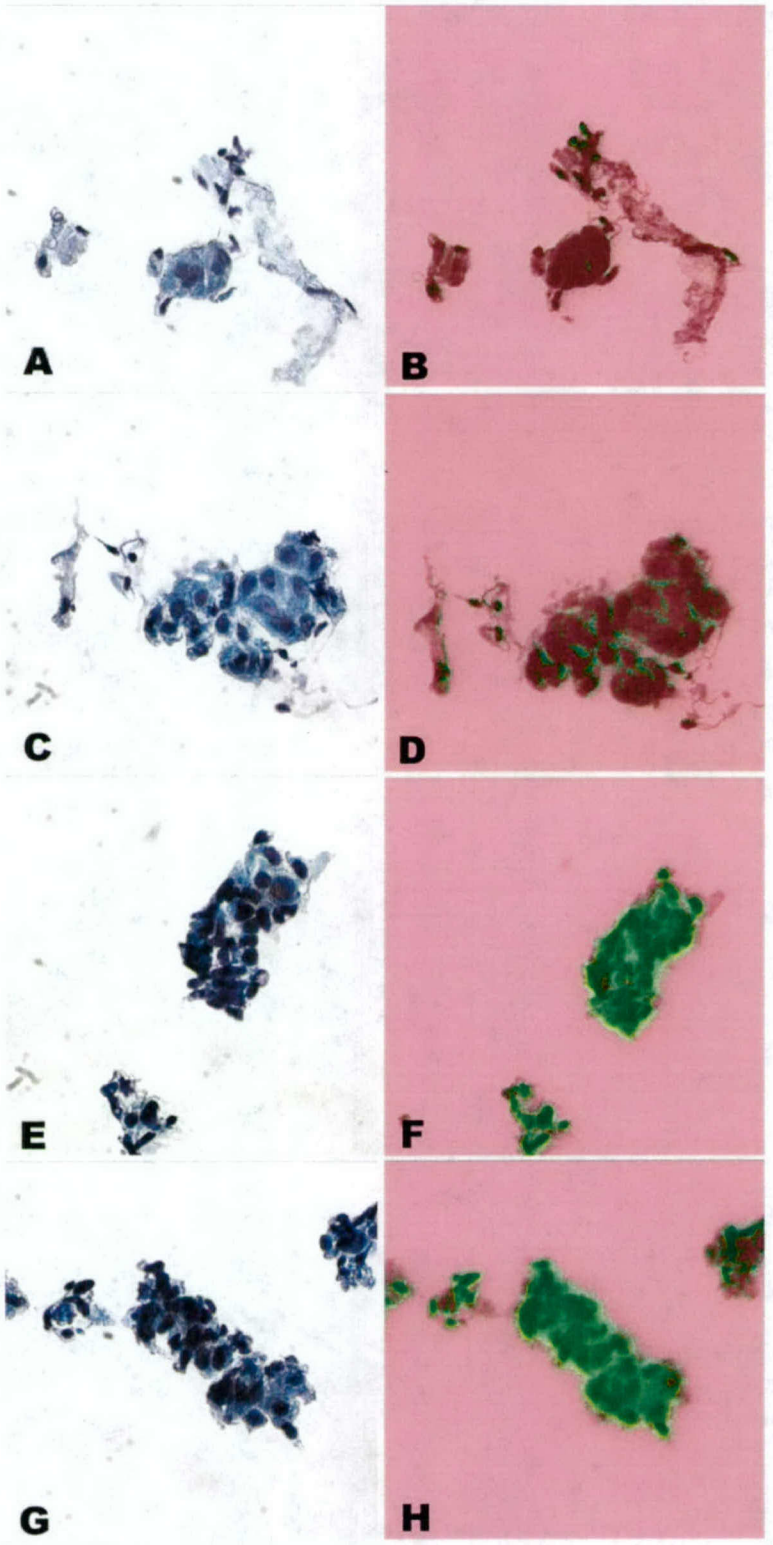

Figure 1a

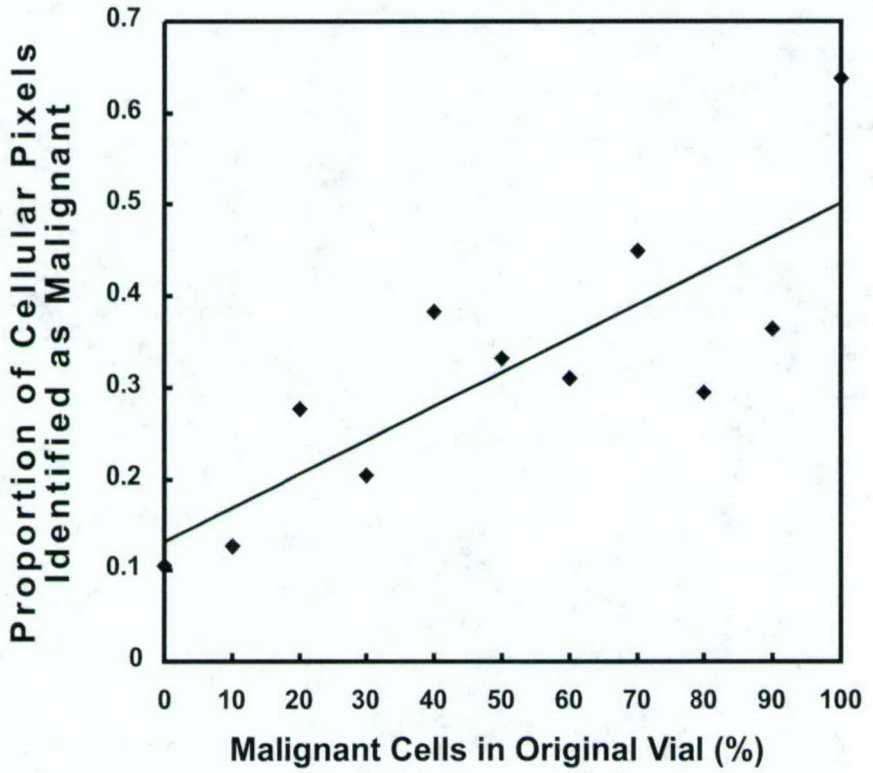
Proportion of Cellular Pixels Identified as Malignant
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 10 20 30 40 50 60 70 80 90 100
Malignant Cells in Original Vial (%)

Figure 1b

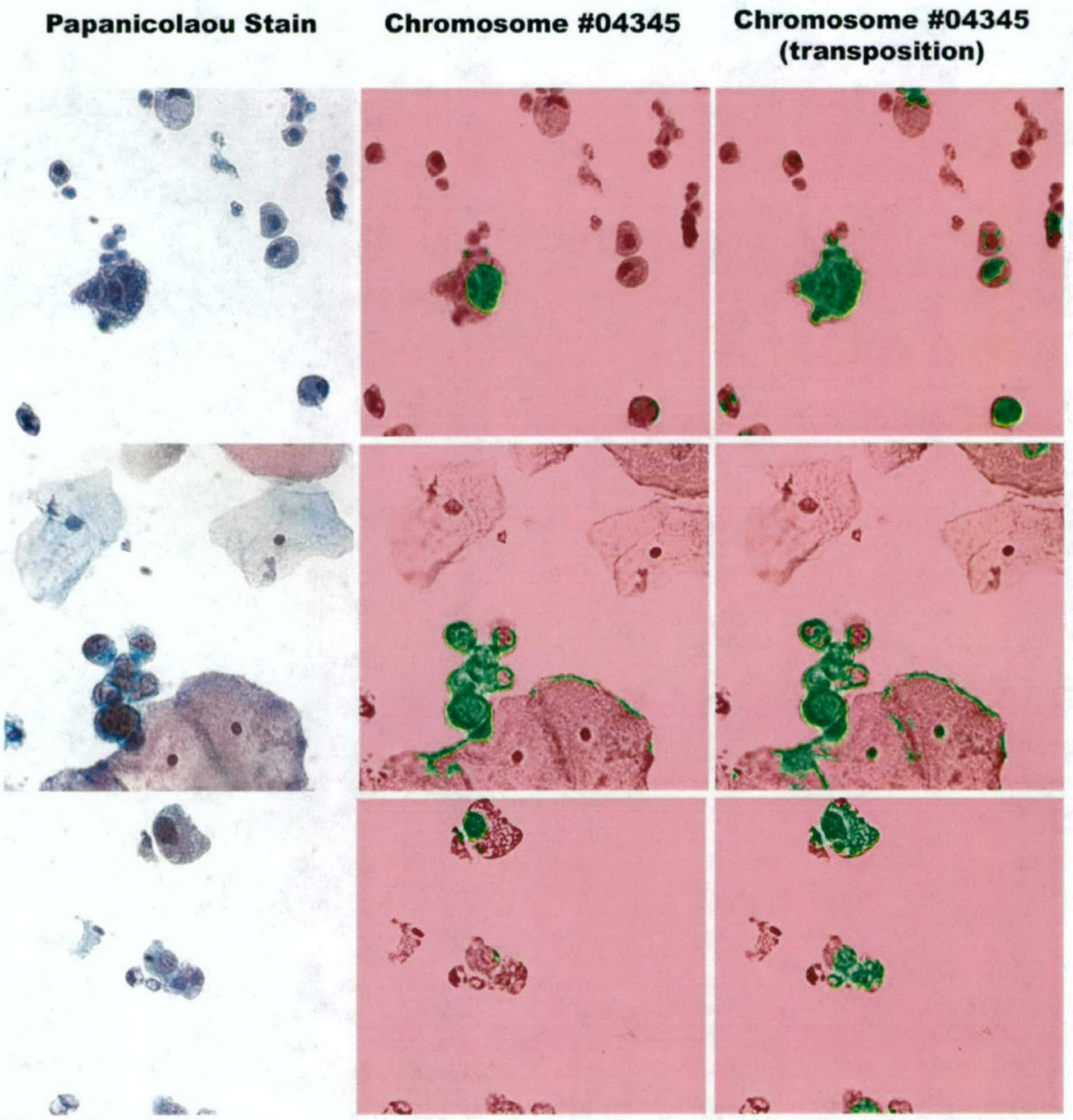
Papanicolaou Stain
Chromosome #04345
Chromosome #04345
(transposition)

Figure 2

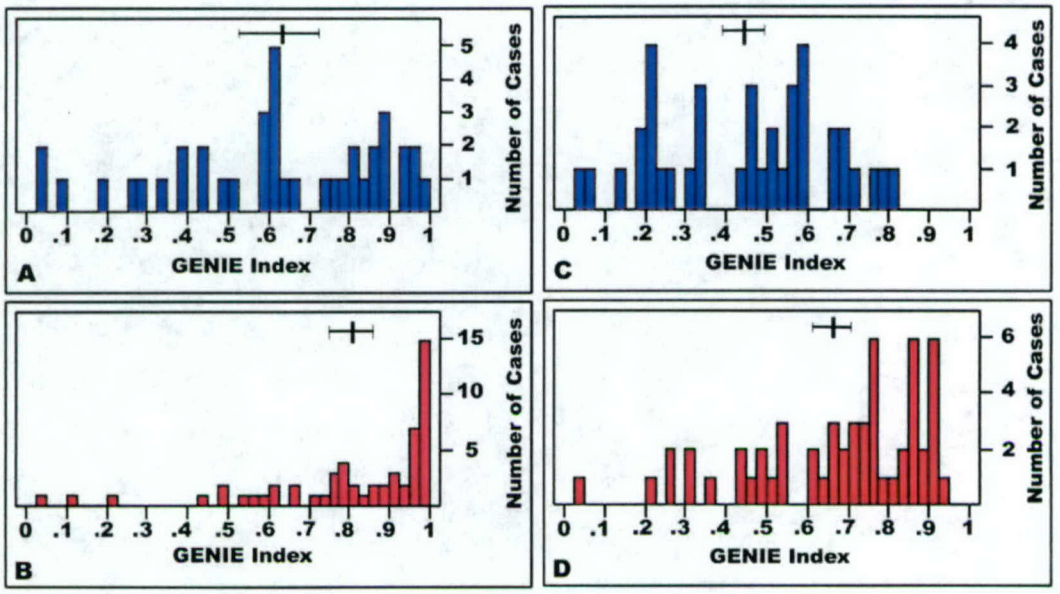

Figure 3

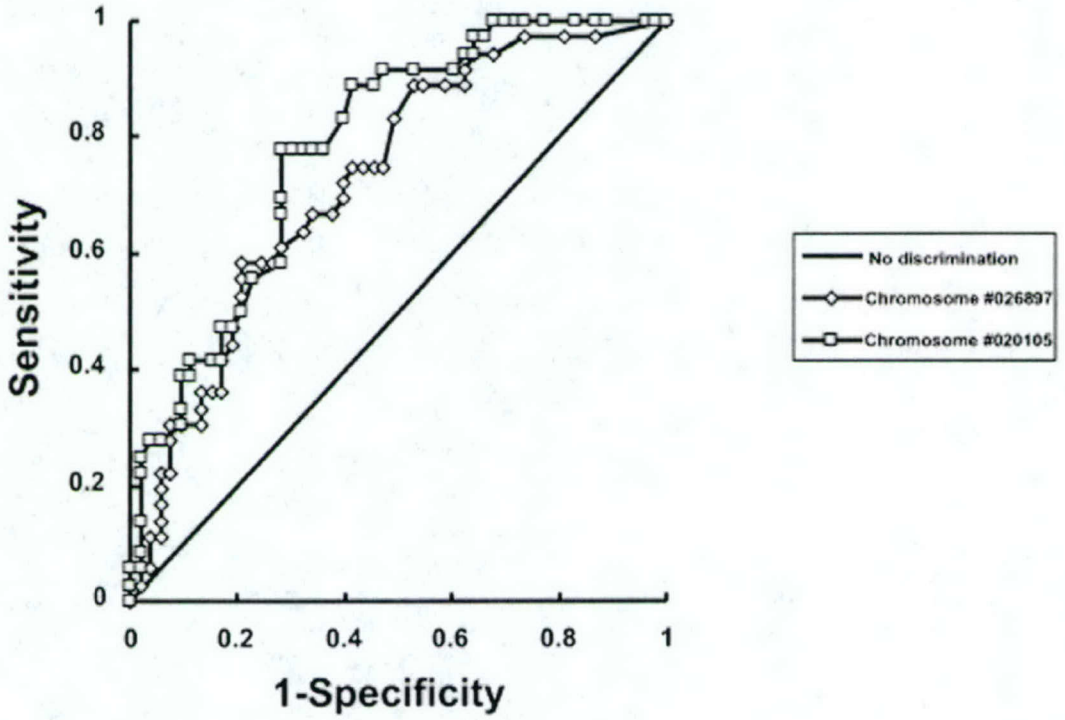

Figure 4

Supplemental Material:

Supplemental Table 1: Details on GENIE Chromosomes:

Chromosome Number (from text)	Associated GENIE algorithm (gene) set
id # 004345	[ASF_CLOP rD5 wS1 2 1][SUBP rD29 rD24 wS2][SQRT rD12 wS0].
#background_092603_000004	[RANGE rD24 wS2 2 1][MIN rS2 rD6 wS0][ADDP rD19 rD8 wS1][ASF_OPCL rD22 wS2 2 1
id #025867	[ASF_OPCL rD15 wS2 6 1][NEG_TH rS2 wS0 9 3][RANGE rD13 wS1 8 1][ASF_OPCL rS0 wS0 7 1]
id #026897	[QTREG rD26 wS1 wS2 wS0 0.05][SQRT rD13 wS1][SADIST rD4 rD23 rD5 rD21 rD16 rD22 rD0 rD24 rS1 rD7 wS0 –1.18 –1.00 1.38 0.94 –1.21 –1.24 1.23 1.00 0.88 -.095].
#020105	[SANORM rD15 rD17 rD10 rD25 wS1 wS0 wS2][DILATE rD16 wS0 2 7][MIN rS1 rD2 wS1][CLIP_HI rD12 wS2 0.06][SD rS1 wS3 5 0][POS_TH rS2 wS1 10 3]

Chromosome explanation:

The chromosome was composed of three operators aligned in the following order: [ASF_CLOP rD5 wS1 2 1][SUBP rD29 rD24 wS2][SQRT rD12 wS0]. ASF_CLOP is an alternating sequential filter (close-open), which basically performs an open-close operation on data plane D5 (470 nm) at increasing sizes using a circular structuring element, writing the result in scratch plane S1. SUBP is a subtract plane operator that subtracts data plane D24 (660 nm) from data plane D29 (700 nm). SQRT performs pixel-by-pixel square root calculation in data plane D12 (540 nm) and writes the result in scratch S0. The results from each scratch plane are added into a feature plane and a threshold value is calculated.

The chromosome generated, #background_092603_000004 had the following sequence: [RANGE rD24 wS2 2 1][MIN rS2 rD6 wS0][ADDP rD19 rD8 wS1][ASF_OPCL rD22 wS2 2 1]. Based on this sequence, the algorithm performs a combination of pixel-neighborhood operations (RANGE; ASF_OPCL), addition between two data planes (ADDP) and a logical operation (MIN).

Chromosome #026897 for atypical urothelial cells had the following sequence: [QTREG rD26 wS1 wS2 wS0 0.05][SQRT rD13 wS1][SADIST rD4 rD23 rD5 rD21 rD16 rD22 rD0 rD24 rS1 rD7 wS0 –1.18 –1.00 1.38 0.94 –1.21 –1.24 1.23 1.00 0.88 -.095]. The first gene (QTREG) performs statistical analysis in relation to region size while the second gene (SQRT) is a basic mathematical operation (square root). The third gene (SADIST) performs spectral angle calculations between neighborhood regions.

Chromosome #020105. for atypical urothelial cells, second attempt larger training set. This is a relatively complex chromosome that performs spectral analysis (SANORM), pixel neighborhood operations (DILATE, SD: Standard Deviation, POS_TH: Positive Top-Hat), threshold operations (CLIP_HI: Clip-High), and a logical pixel-wise minimum (MIN).